



Universidad del Norte Santo Tomás de Aquino

Facultad de Ingeniería

Bioingeniería

Bioinformática

Proyecto Final:

**Análisis integrado de coexpresión y redes de interacción proteica
centrado en SNCA para la identificación de genes candidatos asociados a
procesos neurodegenerativos**

BRIEGA, Ariana Martina

Docente:

Esp. Lic. Ignacio Labari

11 de noviembre de 2025

Tabla de contenido

1.	Introducción	5
2.	Marco Teórico	6
2.1.	Enfermedad de Parkinson [EP]	6
2.1.1.	Características Clínicas	6
2.1.2.	Aspectos moleculares y patogénicos.....	6
2.1.3.	Etiología y factores de riesgo.....	6
2.1.4.	Tratamientos actuales	6
2.1.5.	Líneas actuales de investigación	7
2.1.6.	Importancia del estudio de redes en el contexto del Parkinson.....	7
2.2.	Co – expresión génica.....	7
2.2.1.	Fundamentos del análisis de coexpresión	8
2.2.2.	Módulos de coexpresión	8
2.2.3.	Importancia y aplicaciones.....	9
2.3.	Redes Biologicas.....	9
2.3.1.	Estructura básica de una red.....	9
2.3.2.	Tipos principales de redes biológicas.....	9
2.3.3.	Métricas topológicas y su interpretación biológica.....	10
2.3.4.	Nodos centrales (hubs).....	11
2.4.	Bases de datos y fuentes de información	11
2.4.1.	Gene Expression Omnibus (GEO)	11
2.4.1.	STRING Database.....	12
2.4.2.	Cytoscape	12
2.4.3.	Galaxy Project.....	12
2.4.1.	Otras herramientas bioinformáticas complementarias	13
2.5.	Herramientas estadísticas y computacionales aplicadas	13
2.5.1.	Análisis de Componentes Principales (PCA).....	13
2.5.2.	K-Means Clustering	13
2.5.3.	Clustering jerárquico y heatmaps.....	13
2.5.4.	Análisis de varianza y selección de genes informativos	14
2.5.5.	Redes y métricas de centralidad (Degree, Betweenness)	14
3.	Metodología	14
3.1.	Conjunto de datos	14

3.1.1.	GSE7621	14
3.1.2.	Motivo de elección del dataset	14
3.1.3.	Estructura y formato de los archivos utilizados	15
3.1.4.	Preparación previa del dataset en Galaxy	15
3.2.	Preparación y procesamiento de los datos	15
3.2.1.	Limpieza y control de calidad	15
3.2.2.	Estructuración y alineamiento con metadatos	16
3.2.3.	Normalización de los valores de expresión	16
3.2.4.	Resultados del preprocesamiento	16
3.3.	Análisis con PCA	17
3.3.1.	Aplicación	17
3.3.2.	Visualización de la distribución de las muestras	17
3.4.	Clustering de Muestras	17
3.4.1.	K – means	17
3.4.2.	Jerárquico	17
3.5.	Selección de genes informativos	18
3.5.1.	3.5.1. Cálculo de la variabilidad génica	18
3.5.2.	Selección de los genes más variables	18
3.6.	Construcción de la red de coexpresión génica	19
3.6.1.	Cálculo de correlaciones y filtrado	19
3.6.2.	Detección de módulos	19
3.7.	Análisis de enriquecimiento funcional por módulo	19
3.7.1.	Fundamento del análisis funcional	19
3.7.2.	Herramientas empleadas	20
3.7.3.	Tipos de procesos identificados	20
3.8.	Integración con la red centrada en SNCA	20
3.8.1.	Obtención de la red de SNCA	20
3.8.2.	Superposición con los módulos de coexpresión	21
3.9.	Visualización e integración final en Cytoscape	21
3.9.1.	Importación y estructura de la red	21
3.9.2.	Cálculo de medidas topológicas	21
3.9.3.	Interpretación y valor del enfoque visual	21
3.10.	Priorización de genes candidatos	22

3.10.1.	Criterios de priorización	22
3.10.1.	Asignación del puntaje total	22
3.11.	De análisis bioinformático a propuestas biológicas.....	23
4.	Resultados	23
4.1.	Preparación y estructura final del dataset	23
4.2.	Análisis y distribución de las muestras usando PCA.....	24
4.3.	K – means	25
4.4.	Selección de genes informativos y red de coexpresion.....	27
4.5.	Análisis funcional de los módulos de coexpresion	30
4.5.1.	Modulo M1: Núcleo funcional energético	30
4.5.2.	M2: Núcleo funcional de traducción y de actividades ribosomales	31
4.5.3.	M3: Núcleo funcional de dinámica del citoesqueleto	33
4.5.4.	M4: Núcleo funcional mitocondrial respiratorio.....	34
4.5.5.	M5: Núcleo funcional sináptico y de señalización neuronal.....	35
4.5.6.	M6: transporte y metabolismo de neurotransmisores.....	37
4.5.7.	M7: submodulo sináptico con evidencia limitada.....	38
4.5.8.	M8 y M9.....	39
4.6.	Resultados finales: integración global de la red de coexpresión.....	39
4.6.1.	Integración funcional de los módulos M1–M8	39
4.6.2.	Integración M1-M6 con SNCA en STRING	40
4.6.3.	Integración centrada en SNCA en Cytoscape	41
4.6.4.	Reasignación de módulos en la red integrada	42
4.6.5.	Identificación de hubs	43
4.7.	Priorización de genes candidatos	44
4.8.	De análisis bioinformático a propuestas biológicas	46
5.	Conclusión.....	46
6.	Bibliografía	47

1. Introducción

La comprensión de los mecanismos moleculares que subyacen a las enfermedades neurodegenerativas representa uno de los mayores desafíos actuales de la biología y la medicina. Dentro de este grupo de patologías, la enfermedad de Parkinson ocupa un lugar central por su elevada prevalencia y su compleja etiología multifactorial, que involucra tanto factores genéticos como ambientales. En este contexto, el gen SNCA, que codifica para la α -sinucleína, ha sido identificado como un elemento clave en la fisiopatología de múltiples desórdenes neurodegenerativos conocidos como sinucleinopatías, entre ellos el Parkinson, la demencia con cuerpos de Lewy y la atrofia multisistémica.

La α -sinucleína es una proteína neuronal abundante, localizada principalmente en las terminaciones presinápticas, donde participa en la regulación del tráfico vesicular, la liberación de neurotransmisores y la organización del citoesqueleto. Sin embargo, su sobreexpresión, malplegamiento o agregación anómala puede desencadenar alteraciones celulares que contribuyen a la disfunción sináptica y a la muerte neuronal. A pesar de los avances en el estudio de SNCA, aún no se comprende completamente cómo se integra dentro de las redes moleculares que controlan los procesos neuronales y cómo sus interacciones con otros genes y proteínas pueden modular su función o disfunción.

En las últimas décadas, el enfoque de análisis de redes ha adquirido gran relevancia para estudiar sistemas biológicos complejos. A diferencia de los métodos tradicionales centrados en genes individuales, las redes de coexpresión e interacción proteína-proteína (PPI) permiten capturar relaciones globales, identificar módulos funcionales y detectar nodos centrales (hubs) que podrían desempeñar roles regulatorios o estructurales críticos. Estas aproximaciones son particularmente útiles para caracterizar el contexto molecular de genes como SNCA, cuya función depende en gran medida de su conectividad con otras proteínas.

El presente trabajo propone un análisis integrativo que combina información de coexpresión génica y datos de interacción funcional obtenidos de STRING, con el objetivo de caracterizar la posición de SNCA dentro de una red biológica más amplia y priorizar genes candidatos potencialmente asociados a su función o disfunción neuronal. Para ello, se construyó una red a partir de 208 genes asociados a SNCA, se incorporaron interacciones experimentales y predichas, y se aplicaron métricas de centralidad (Degree y Betweenness) mediante la herramienta Cytoscape.

Finalmente, se implementó una estrategia de priorización de candidatos basada en tres criterios: pertenencia a módulos funcionales relevantes, alta conectividad topológica y conexión directa con SNCA. Este enfoque permitió identificar genes con alta probabilidad de participación en procesos sinápticos y neurodegenerativos, aportando una base de datos de interés para futuros análisis funcionales o validaciones experimentales.

2. Marco Teórico

2.1. *Enfermedad de Parkinson [EP]*

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo progresivo que afecta principalmente al sistema nervioso central, caracterizado por la pérdida selectiva de neuronas dopaminérgicas en la *sustancia negra pars compacta*, una región del mesencéfalo encargada de modular el control motor. Esta pérdida conduce a una disminución en los niveles de dopamina en el estriado, generando un desequilibrio en los circuitos neuronales que regulan el movimiento.

2.1.1. *Características Clínicas*

Desde el punto de vista clínico, la EP se manifiesta a través de síntomas motores clásicos, como temblor en reposo, rigidez muscular, bradicinesia (lentitud de movimientos) y alteraciones posturales, acompañados frecuentemente de síntomas no motores, como trastornos del sueño, alteraciones cognitivas, depresión y disfunción autonómica. Estos síntomas suelen progresar lentamente a lo largo de los años, afectando de manera significativa la calidad de vida del paciente.

2.1.2. *Aspectos moleculares y patogénicos*

A nivel molecular, uno de los eventos más característicos de la enfermedad es la acumulación anómala de la proteína α -sinucleína, codificada por el gen SNCA, dentro de las neuronas. Esta proteína, que normalmente participa en la regulación del tráfico vesicular y la liberación de neurotransmisores en las terminales presinápticas, puede sufrir malplegamiento y agregarse en estructuras fibrilares conocidas como cuerpos de Lewy, un sello histopatológico distintivo de la enfermedad.

La acumulación de α -sinucleína alterada desencadena estrés del retículo endoplasmático, disfunción mitocondrial y estrés oxidativo, promoviendo un ciclo de daño neuronal progresivo. Además, se ha observado que estos agregados pueden propagarse entre neuronas, lo que sugiere un mecanismo de tipo “priónico” en la diseminación de la patología.

2.1.3. *Etiología y factores de riesgo*

Aunque la mayoría de los casos son esporádicos, se han identificado formas familiares de la enfermedad asociadas a mutaciones en genes específicos, entre ellos SNCA, LRRK2, PARK2 (Parkin), PINK1 y DJ-1.

Los factores ambientales también desempeñan un papel importante: la exposición prolongada a pesticidas, solventes y metales pesados ha sido relacionada con un mayor riesgo de desarrollar la enfermedad, mientras que hábitos como el consumo moderado de cafeína o el tabaquismo parecen tener un efecto protector, aunque los mecanismos detrás de estas asociaciones aún se investigan.

2.1.4. *Tratamientos actuales*

Actualmente, el tratamiento del Parkinson se centra en aliviar los síntomas, ya que no existe una cura definitiva.

El enfoque farmacológico más utilizado es la terapia dopaminérgica, principalmente con levodopa, un precursor de la dopamina que atraviesa la barrera hematoencefálica. Esta se combina frecuentemente con inhibidores de la dopa descarboxilasa o agonistas dopaminérgicos para prolongar su efecto y reducir efectos secundarios.

En casos más avanzados, se puede recurrir a tratamientos quirúrgicos como la estimulación cerebral profunda (DBS), que modula la actividad eléctrica de los ganglios basales.

Sin embargo, estas terapias son sintomáticas y no detienen la progresión de la enfermedad, lo que impulsa la búsqueda de nuevas estrategias terapéuticas dirigidas a los mecanismos celulares subyacentes, como la autofagia-lisosoma, la proteostasis y la biogénesis mitocondrial.

2.1.5. *Líneas actuales de investigación*

Las investigaciones recientes se orientan hacia la identificación de biomarcadores tempranos y dianas moleculares que permitan desarrollar terapias modificadoras de la enfermedad. Entre ellas se destacan:

- Moduladores de la expresión o agregación de α -sinucleína, como anticuerpos monoclonales y moléculas chaperonas.
- Activadores de la autofagia (por ejemplo, TFEB, AMPK-mTOR), que promueven la degradación de agregados proteicos.
- Compuestos antioxidantes y moduladores de la función mitocondrial (PGC-1 α , Nrf2).
- Terapias génicas y celulares, dirigidas a restaurar la función dopaminérgica o prevenir la muerte neuronal.

2.1.6. *Importancia del estudio de redes en el contexto del Parkinson*

Dado que la EP no se debe a un único gen o evento molecular, sino a una red compleja de interacciones, el enfoque de análisis de redes permite integrar y comprender las relaciones funcionales entre genes y proteínas involucradas en la neurodegeneración.

En particular, estudiar la posición de SNCA dentro de una red de coexpresión y de interacciones proteicas brinda una perspectiva más completa sobre cómo su disfunción puede alterar procesos celulares interconectados, como el transporte vesicular, la homeostasis mitocondrial y la respuesta al estrés proteico.

2.2. *Co – expresión génica*

La coexpresión génica es un enfoque analítico dentro de la biología de sistemas que permite identificar genes que presentan patrones de expresión similares a través de múltiples condiciones biológicas, muestras o tejidos.

En otras palabras, cuando dos o más genes muestran niveles de expresión que varían de forma coordinada (por ejemplo, aumentan o disminuyen juntos en distintas muestras), se considera que están coexpresados.

Esta relación no necesariamente implica interacción física directa, pero sí sugiere una posible conexión funcional o regulatoria, por ejemplo, que los genes estén controlados por el mismo factor de transcripción, que participen en una misma vía metabólica o que formen parte de un mismo complejo proteico.

2.2.1. *Fundamentos del análisis de coexpresión*

El análisis de coexpresión se basa en la correlación estadística entre los niveles de expresión génica.

A partir de una matriz de expresión (donde las filas representan genes y las columnas representan muestras o condiciones), se calcula un coeficiente de correlación (por ejemplo, de Pearson o Spearman) entre cada par de genes.

El resultado es una matriz de correlaciones que refleja la fuerza y dirección de la relación entre todos los pares de genes del conjunto.

A partir de esta información, es posible construir una red de coexpresión, donde:

- cada nodo representa un gen, y
- cada arista (edge) representa una correlación significativa entre dos genes.

Estas redes son útiles para descubrir grupos o módulos de genes que comparten funciones biológicas o que podrían estar co-regulados.

2.2.2. *Módulos de coexpresión*

Uno de los aspectos más valiosos de la coexpresión es la posibilidad de identificar módulos: subconjuntos de genes altamente interconectados entre sí, pero débilmente conectados con el resto de la red.

Cada módulo suele corresponder a un proceso biológico específico, como “respuesta al estrés oxidativo”, “transporte vesicular” o “actividad sináptica”.

La identificación de estos módulos permite:

- Reducir la complejidad del análisis genómico (de miles de genes a grupos funcionales).
- Detectar patrones de disfunción coordinada, en lugar de efectos aislados.
- Priorizar genes candidatos que podrían estar implicados en una misma vía patológica.

En el contexto del Parkinson, este enfoque es especialmente útil, ya que la enfermedad involucra alteraciones simultáneas en múltiples procesos: proteostasis, función mitocondrial, autofagia, metabolismo energético y señalización sináptica.

Analizar cómo los genes se coexpresan permite observar qué grupos funcionales están afectados en conjunto.

2.2.3. *Importancia y aplicaciones*

El análisis de coexpresión tiene múltiples aplicaciones:

- Descubrimiento de biomarcadores: genes con patrones de expresión comunes a un estado patológico.
- Predicción de función génica: si un gen desconocido se coexpresa con genes de función conocida, puede inferirse su rol.
- Integración con redes de interacción proteica (PPI): permite vincular la co-regulación con la interacción física entre proteínas.

En investigaciones neurobiológicas, esta metodología ayuda a identificar programas de expresión neuronales o gliales que podrían alterarse durante la neurodegeneración.

2.3. *Redes Biológicas*

En biología, muchos procesos celulares no dependen de la acción de un solo gen o proteína, sino de interacciones dinámicas entre múltiples componentes. Estas relaciones pueden representarse mediante redes biológicas, un enfoque que permite visualizar y analizar sistemas complejos a través de una estructura matemática compuesta por nodos (genes, proteínas, metabolitos) y aristas (las relaciones o interacciones entre ellos).

Las redes biológicas constituyen una herramienta central en la biología de sistemas, ya que integran información molecular, funcional y regulatoria, ofreciendo una visión más global del comportamiento celular.

2.3.1. *Estructura básica de una red*

- Nodos: representan los elementos individuales del sistema (genes, proteínas o metabolitos).
- Aristas (edges): representan las interacciones o relaciones entre nodos. Estas pueden ser:
 - Físicas, como interacciones proteína-proteína (PPI).
 - Funcionales, como correlaciones de expresión (coexpresión).
 - Regulatorias, como interacciones entre un factor de transcripción y sus genes blanco.

Las redes pueden ser dirigidas (cuando la relación tiene un sentido, por ejemplo, una regulación génica) o no dirigidas (cuando la relación es bidireccional, como la coexpresión). Además, pueden ser ponderadas, asignando un valor numérico a cada arista que indica la fuerza o confianza de la interacción (por ejemplo, un coeficiente de correlación o un score experimental).

2.3.2. *Tipos principales de redes biológicas*

- Redes de coexpresión génica:
 - Conectan genes con patrones de expresión similares.
 - Revelan grupos co-regulados o vías compartidas.
- Redes de interacción proteína-proteína (PPI):
 - Representan interacciones físicas o funcionales entre proteínas.
 - Permiten explorar cómo las alteraciones en una proteína pueden afectar a otras dentro de una vía biológica.
- Redes reguladoras:
 - Muestran cómo los factores de transcripción controlan la expresión de sus genes blanco.
- Redes metabólicas:
 - Describen cómo los metabolitos se transforman entre sí mediante reacciones enzimáticas.

2.3.3. Métricas topológicas y su interpretación biológica

El análisis de redes biológicas incluye el cálculo de métricas de centralidad, que cuantifican la posición e importancia relativa de cada nodo dentro del sistema. Estas medidas son esenciales para identificar genes clave (hubs), mediadores o conectores funcionales.

Métrica	Que mide	Interpretación biológica
Degree (grado)	Numero de conexiones directas de un nodo	Indica que tan conectado esta un gen o proteína dentro de la red. Los nodos con alto grado son llamados hubs y suelen cumplir funciones esenciales
Betweenness centrality	Cuantos caminos mas cortos pasan por ese nodo	Identifica nodos que actúan como puentes o mediadores entre módulos funcionales distintos.
Closeness centrality	Inverso de la distancia promedio a todos los demás nodos	Identifica la capacidad de un nodo para influir rápidamente en toda la red
Modularidad	Agrupamiento o densidad interna de conexiones dentro de subredes	Permite identificar módulos funcionales o comunidades de genes con roles biológicos similares

Tabla 1. Métricas de redes biológicas

Estas métricas no solo describen la estructura de la red, sino que permiten inferir roles biológicos. Por ejemplo, en una red de interacción proteica:

- Un nodo con alto Degree puede representar una proteína estructural o reguladora central.
- Un nodo con alta Betweenness puede ser una proteína señalizadora que conecta rutas metabólicas distintas.

2.3.4. *Nodos centrales (hubs)*

Los hubs son nodos con un número de conexiones significativamente superior al promedio. En las redes biológicas, suelen corresponder a genes o proteínas esenciales para la supervivencia celular.

Su eliminación o disfunción puede tener efectos drásticos, ya que afectan a múltiples procesos simultáneamente.

Existen dos tipos principales:

- Hubs locales: muy conectados dentro de un módulo funcional.
- Hubs globales: conectan diferentes módulos o vías biológicas, actuando como puentes entre sistemas.

2.4. *Bases de datos y fuentes de información*

El avance de la biología molecular y la bioinformática ha generado una enorme cantidad de datos experimentales, cuyo análisis sistemático requiere el uso de bases de datos públicas y recursos computacionales especializados. Estas herramientas permiten integrar información genómica, transcriptómica y proteómica, favoreciendo una comprensión global de los sistemas biológicos.

En el contexto del estudio de enfermedades complejas como el Parkinson, las bases de datos son esenciales para:

- Acceder a datos experimentales generados por diferentes grupos de investigación,
- Comparar perfiles de expresión entre condiciones biológicas (por ejemplo, control vs. enfermedad),
- Relacionar esos patrones con redes funcionales de proteínas, vías metabólicas y procesos celulares.

A continuación, se presentan las principales fuentes conceptuales utilizadas en este tipo de análisis.

2.4.1. *Gene Expression Omnibus (GEO)*

GEO es una base de datos pública del National Center for Biotechnology Information (NCBI) que almacena resultados de experimentos de transcriptómica, tanto de microarreglos (microarrays) como de secuenciación de ARN (RNA-seq).

Cada estudio depositado en GEO contiene información sobre:

- El diseño experimental (tipo de tejido, condición biológica, número de muestras),
- Las matrices de expresión génica resultantes,
- Y los metadatos que describen las variables clínicas o experimentales asociadas.

Estos datos se agrupan en Series (GSE), que representan conjuntos de muestras relacionadas, y Plataformas (GPL), que indican el tipo de chip o tecnología empleada.

El valor de GEO radica en que permite reutilizar datos experimentales originales para realizar nuevos análisis, aplicar métodos estadísticos o exploraciones bioinformáticas distintas, favoreciendo la reproducibilidad científica y el descubrimiento secundario.

En el caso del Parkinson, GEO contiene numerosos datasets de tejido cerebral, líneas celulares y modelos animales, lo que la convierte en una fuente fundamental para estudiar patrones de expresión diferencial y coexpresión génica.

2.4.1. *STRING Database*

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) es una base de datos que integra información sobre interacciones proteína-proteína (PPI).

Cada interacción en STRING posee un “confidence score” (valor de confianza), que indica la probabilidad de que la relación sea biológicamente real.

No solo permite visualizar redes PPI, sino también realizar análisis de enriquecimiento funcional, como:

- Gene Ontology (GO): clasificación por función biológica, componente celular o proceso molecular.
- KEGG y Reactome: mapeo de vías metabólicas o señalización.
- Clusters funcionales: agrupamiento de proteínas en módulos coherentes.

2.4.2. *Cytoscape*

Cytoscape es una plataforma de software libre para la visualización y análisis de redes biológicas.

Permite integrar distintos tipos de datos (por ejemplo, resultados de expresión génica, métricas de centralidad o enriquecimiento funcional) dentro de una misma red, facilitando la interpretación visual y cuantitativa de los sistemas moleculares.

Es ampliamente utilizado para explorar jerarquías y relaciones entre genes o proteínas, identificar nodos clave (hubs) y comunicar los resultados de manera visual y comprensible.

2.4.3. *Galaxy Project*

Galaxy es una plataforma web de análisis bioinformático que permite ejecutar flujos de trabajo (workflows) complejos sin necesidad de programación.

Ofrece acceso a una amplia gama de herramientas para el procesamiento, análisis estadístico y visualización de datos biológicos, incluyendo módulos de expresión génica, clustering, enriquecimiento funcional y generación de redes.

2.4.1. Otras herramientas bioinformáticas complementarias

- Pandas, Scikit-learn y Matplotlib (Python): librerías esenciales para la manipulación de datos, normalización, reducción de dimensionalidad (PCA) y visualización de resultados.

2.5. Herramientas estadísticas y computacionales aplicadas

El análisis de expresión génica a gran escala requiere métodos estadísticos y computacionales que permitan reducir la complejidad de los datos, identificar patrones y agrupar muestras o genes con comportamientos similares.

En este trabajo se emplearon diversas técnicas complementarias, tanto exploratorias como de agrupamiento, con el objetivo de interpretar la estructura subyacente de los datos transcriptómicos.

2.5.1. Análisis de Componentes Principales (PCA)

El PCA (Principal Component Analysis) es una técnica de reducción de dimensionalidad que transforma miles de variables (genes) en un número menor de componentes principales, que resumen la mayor parte de la variabilidad del conjunto de datos.

Permite:

- Detectar patrones globales de variación.
- Evaluar si las muestras tienden a agruparse según su condición (por ejemplo, Parkinson vs. control).
- Visualizar posibles outliers o diferencias entre grupos.

2.5.2. K-Means Clustering

El método de K-Means agrupa los datos en un número predefinido de clusters (k), de modo que las muestras dentro de un mismo grupo sean más similares entre sí que respecto a las de otros grupos.

Es una técnica no supervisada, ideal para descubrimiento exploratorio cuando no se conocen de antemano los grupos biológicos.

2.5.3. Clustering jerárquico y heatmaps

El clustering jerárquico organiza los genes o muestras en forma de árbol (dendrograma), según su similitud.

Su visualización mediante heatmaps (mapas de calor) permite observar grupos de genes con comportamientos coordinados.

2.5.4. *Análisis de varianza y selección de genes informativos*

Dado que no todos los genes contribuyen de igual manera a la variabilidad del sistema, se calcularon medidas de varianza por gen para seleccionar los más informativos. Los genes con mayor varianza suelen corresponder a los que cambian significativamente entre condiciones, por lo que su inclusión mejora la sensibilidad de los análisis posteriores (como coexpresión o clustering).

2.5.5. *Redes y métricas de centralidad (Degree, Betweenness)*

En la fase de análisis de redes, se aplicaron métricas de centralidad para cuantificar la relevancia de cada gen dentro de la red de coexpresión.

3. Metodología

3.1. *Conjunto de datos*

Para el presente trabajo se utilizó un conjunto de datos público disponible en la base Gene Expression Omnibus (GEO), dependiente del *National Center for Biotechnology Information (NCBI)*.

El dataset seleccionado fue GSE7621, titulado “*Gene expression profiling of Parkinson’s disease substantia nigra*”. Este estudio forma parte de una línea de investigación destinada a caracterizar los cambios en la expresión génica en la sustancia negra del cerebro, una región crítica para el control del movimiento y una de las más afectadas en la enfermedad de Parkinson.

3.1.1. *GSE7621*

El estudio GSE7621 fue conducido por Lesnick et al. (2007) y depositado en GEO con el objetivo de comparar la expresión génica en tejido cerebral post mortem de individuos diagnosticados con enfermedad de Parkinson (PD) y sujetos control sin patología neurológica. Las muestras fueron tomadas de la *sustancia negra pars compacta*, donde ocurre la degeneración dopaminérgica característica del Parkinson.

El conjunto de datos incluye:

- 16 muestras en total: 9 correspondientes a pacientes con enfermedad de Parkinson y 7 controles sanos.
- Cada muestra cuenta con un archivo de expresión individual y un conjunto de metadatos que incluyen el tipo de tejido, la condición clínica y el código identificador de muestra.

3.1.2. *Motivo de elección del dataset*

- Relevancia biológica: analiza directamente la región cerebral más vinculada con la patogénesis del Parkinson.
- Disponibilidad de metadatos claros: permite diferenciar entre las condiciones *Control* y *PD* con precisión.
- Calidad y accesibilidad: el estudio está ampliamente citado y presenta datos bien curados, compatibles con herramientas de análisis actuales.
- Compatibilidad con redes de coexpresión: al tratarse de un microarray con amplia cobertura génica, permite explorar correlaciones robustas entre genes.

3.1.3. *Estructura y formato de los archivos utilizados*

Se descargaron dos archivos principales:

- Una matriz de expresión génica (“Galaxy26-[Matriz por gen definitiva].tabular”), donde cada fila representa un gen y cada columna una muestra.
- Un archivo de metadatos (“metadata_GSE7621.tsv”) que contiene el identificador de cada muestra (sample ID) y su condición (Control o PD).

Antes del análisis, la matriz fue transpuesta (Galaxy48-[Matriz por gen transpuesta].tabular) para adoptar el formato convencional de análisis bioinformático, donde las filas representan las muestras y las columnas los genes. Esto permitió emparejarla con los metadatos y aplicar posteriormente los métodos estadísticos y de clustering.

3.1.4. *Preparación previa del dataset en Galaxy*

Previo al análisis en Python, se realizó una etapa inicial de preparación de los datos en la plataforma Galaxy, un entorno bioinformático de uso libre que permite ejecutar herramientas de análisis sin necesidad de programación.

El propósito de esta fase fue organizar, limpiar y estructurar la matriz de expresión génica descargada del repositorio GEO para asegurar su compatibilidad con los análisis estadísticos posteriores.

3.2. *Preparación y procesamiento de los datos*

Una vez obtenida y transpuesta la matriz de expresión génica en Galaxy, el siguiente paso consistió en su preprocesamiento dentro del entorno Python, con el objetivo de garantizar la calidad y homogeneidad de los datos antes de aplicar los métodos estadísticos y de agrupamiento.

3.2.1. *Limpieza y control de calidad*

Se cargaron dos archivos principales:

- La matriz de expresión génica, en la que cada columna representa un gen y cada fila una muestra.

- El archivo de metadatos, que contiene la condición biológica de cada muestra (Control o PD).

El primer paso fue eliminar todas aquellas filas que no tuvieran símbolo de gen asignado (valores “NA” o “.”), ya que estos registros no aportan información biológica útil. Posteriormente, se verificó que los valores contenidos fueran numéricos, reemplazando los infinitos por valores nulos y eliminando aquellas columnas (genes) con datos faltantes o sin variación (varianza igual a cero).

Este proceso redujo el ruido técnico y permitió conservar únicamente genes con expresión medible y variable entre las muestras.

3.2.2. Estructuración y alineamiento con metadatos

Para asegurar la correspondencia exacta entre los datos de expresión y los metadatos, se ordenaron las filas de la matriz de acuerdo con el orden de las muestras especificado en el archivo *metadata_GSE7621.tsv*.

Este paso es fundamental para evitar errores en los análisis supervisados y no supervisados posteriores, garantizando que cada perfil de expresión esté correctamente vinculado con su etiqueta (*Control* o *Parkinson*).

3.2.3. Normalización de los valores de expresión

Debido a que los distintos genes pueden presentar escalas de expresión muy diferentes, se aplicó una normalización tipo z-score mediante la clase *StandardScaler* de *Scikit-learn*. Esta transformación centra los datos (media cero) y los escala según su desviación estándar, asegurando que todos los genes tengan el mismo peso en los análisis multivariados.

Matemáticamente, la transformación se define como:

$$z = \frac{x - \mu}{\sigma}$$

donde X representa la expresión original, μ la media y σ la desviación estándar del gen.

3.2.4. Resultados del preprocesamiento

Tras la normalización, se obtuvo una matriz estandarizada lista para los análisis exploratorios.

En total, se conservaron todas las muestras del estudio ($n = 16$) y un número reducido de genes filtrados por calidad.

Esta matriz sirvió como base para la aplicación del Análisis de Componentes Principales (PCA) y los métodos de clustering empleados en las etapas siguientes.

3.3. *Análisis con PCA*

El Análisis de Componentes Principales (PCA, por sus siglas en inglés) fue la primera técnica exploratoria aplicada al conjunto de datos ya normalizado.

El objetivo de este procedimiento fue reducir la dimensionalidad del espacio génico (miles de variables) a un número pequeño de componentes que capturen la mayor parte de la variabilidad total del sistema, permitiendo visualizar las relaciones entre muestras y posibles patrones de agrupamiento según la condición clínica (*Control* o *Parkinson*).

3.3.1. *Aplicación*

El análisis se realizó utilizando la clase PCA de la librería *Scikit-learn*, especificando dos componentes principales (PC1 y PC2).

El modelo fue ajustado sobre la matriz de expresión normalizada y los valores proyectados fueron graficados en un plano bidimensional, donde cada punto representa una muestra.

El porcentaje de varianza explicada por cada componente se calculó automáticamente, indicando cuánta información del total de los datos se conserva en la reducción dimensional.

3.3.2. *Visualización de la distribución de las muestras*

Los resultados se representaron mediante un diagrama de dispersión (scatter plot), donde:

- Cada punto corresponde a una muestra biológica.
- El color indica la condición (azul para controles, rojo para pacientes con Parkinson).

Esta representación permitió observar si existía una separación natural entre los grupos basada únicamente en su perfil de expresión génica global.

3.4. *Clustering de Muestras*

Con el objetivo de explorar patrones de similitud global entre las muestras y evaluar si las condiciones biológicas (*Control* y *Parkinson*) podían agruparse de forma no supervisada, se aplicaron diferentes métodos de clustering sobre los datos proyectados por el PCA.

3.4.1. *K – means*

Su aplicación en este estudio tuvo como propósito detectar posibles subgrupos dentro del conjunto de muestras que compartieran perfiles de expresión similares, sin utilizar información previa sobre la condición clínica.

El algoritmo fue ejecutado sobre los componentes principales obtenidos en el PCA, lo que permitió reducir el ruido y aumentar la estabilidad de la agrupación. Finalmente, los resultados se visualizaron mediante gráficos bidimensionales con centroides y contornos de agrupamiento, siguiendo la estructura presentada en las clases prácticas de la materia.

3.4.2. *Jerárquico*

En paralelo, se implementó un análisis de clustering jerárquico, con el fin de comparar su desempeño y robustez frente al método de K-Means. Este enfoque no requiere especificar el número de grupos a priori, sino que construye una jerarquía de similitudes entre las muestras, uniendo las más similares en etapas sucesivas hasta formar un dendrograma.

El análisis se realizó sobre los mismos datos normalizados y reducidos por PCA, utilizando diferentes estrategias de *linkage* (simple, complete y average) para evaluar la consistencia de la estructura jerárquica.

Los resultados se representaron mediante dendrogramas y mapas de calor (heatmaps) que permiten observar visualmente la proximidad entre las muestras en función de su perfil global de expresión.

3.5. *Selección de genes informativos*

Luego de los análisis exploratorios sobre las muestras, el siguiente paso consistió en identificar los genes más informativos del dataset, es decir, aquellos cuya expresión presenta mayor variabilidad entre las condiciones biológicas.

La finalidad de este procedimiento fue reducir la complejidad del análisis y enfocar los estudios de coexpresión en genes que realmente contribuyen a las diferencias observadas, eliminando aquellos con cambios mínimos o constantes.

3.5.1. *Cálculo de la variabilidad génica*

Para cada gen se calculó la varianza de expresión a lo largo de todas las muestras mediante herramientas estadísticas implementadas en Python. Este valor permite estimar el grado de dispersión de la expresión de un gen:

- Los genes con baja varianza son generalmente constitutivos o poco sensibles a cambios fisiológicos.
- Los genes con alta varianza, en cambio, tienden a reflejar procesos biológicos específicos o alteraciones asociadas a la enfermedad.

3.5.2. *Selección de los genes más variables*

A partir del cálculo anterior, se seleccionó un subconjunto de entre 500 y 1000 genes con mayor varianza, denominados genes informativos o “TopVar genes”. Esta estrategia permite mantener un balance entre representatividad biológica y eficiencia computacional.

El conjunto resultante se almacenó en un archivo TopVar_genes.tsv, que sirvió como punto de partida para la construcción de la red de coexpresión.

La selección basada en variabilidad tiene como ventaja que no depende de comparaciones directas ni de pruebas de hipótesis, lo cual evita sesgos derivados de supuestos estadísticos. Además, al enfocarse en los genes que más cambian entre muestras, se incrementa la sensibilidad

de los análisis posteriores para detectar módulos funcionales y vías biológicas relevantes en la enfermedad de Parkinson.

3.6. Construcción de la red de coexpresión génica

Una vez identificados los genes más variables, se procedió a la construcción de una red de coexpresión, cuyo propósito fue analizar cómo se relacionan entre sí los genes a partir de sus patrones de expresión.

Este tipo de análisis permite identificar módulos de genes que actúan coordinadamente, lo que refleja posibles procesos biológicos compartidos o regulaciones comunes.

3.6.1. Cálculo de correlaciones y filtrado

Se calculó la matriz de correlaciones de Pearson entre todos los pares de genes seleccionados.

Posteriormente, se aplicó un umbral de correlación absoluta ≥ 0.7 , conservando únicamente las asociaciones más fuertes y biológicamente relevantes. De esta manera, la red se centró en interacciones robustas, eliminando relaciones débiles que podrían corresponder a ruido técnico.

3.6.2. Detección de módulos

A partir de la red filtrada, se aplicó el algoritmo de detección de comunidades Louvain, una técnica ampliamente usada para descubrir grupos de genes que se conectan más entre sí que con el resto de la red.

Cada comunidad o módulo representa un conjunto de genes con expresión coordinada, potencialmente asociados a un mismo proceso biológico.

Los módulos fueron denominados M1, M2, M3..., y sus componentes se exportaron para su posterior análisis funcional.

3.7. Análisis de enriquecimiento funcional por módulo

Con los módulos de coexpresión ya identificados, se llevó a cabo un análisis de enriquecimiento funcional para determinar qué procesos biológicos, rutas metabólicas o mecanismos moleculares estaban sobrerrepresentados dentro de cada conjunto de genes.

El objetivo de este análisis fue interpretar biológicamente los módulos, vinculando las correlaciones observadas con funciones celulares relevantes en la enfermedad de Parkinson.

3.7.1. Fundamento del análisis funcional

El enriquecimiento funcional consiste en comparar la lista de genes de un módulo con bases de datos de conocimiento biológico (como GEO) identificando aquellos términos que aparecen con mayor frecuencia de lo esperado por azar.

De este modo, es posible inferir qué funciones o rutas biológicas comparten los genes que se coexpresan, ofreciendo una interpretación funcional de los módulos detectados.

3.7.2. *Herramientas empleadas*

El análisis se realizó utilizando la plataforma STRING que permiten ingresar listas de genes y obtener los procesos enriquecidos con sus respectivos valores de significancia estadística (p-ajustada o FDR).

Cada módulo (M1, M2, M3, etc.) fue analizado por separado, empleando como organismo de referencia *Homo sapiens* y estableciendo un umbral de significancia de p-ajustada < 0.05 .

Para cada módulo se generaron salidas en formato tabular (.tsv) y visualizaciones gráficas (gráficos de barras o burbujas), donde se destacaron las principales categorías funcionales enriquecidas.

3.7.3. *Tipos de procesos identificados*

Si bien los resultados específicos se detallan más adelante, el análisis de enriquecimiento permitió observar que diferentes módulos se asocian a distintas funciones celulares, muchas de ellas relevantes para la neurodegeneración, como:

- Procesos de sinapsis y neurotransmisión.
- Proteostasis
- Autofagia y lisosomas.
- Respuesta al estrés oxidativo y función mitocondrial.

Este tipo de análisis es fundamental para pasar de patrones estadísticos a interpretaciones biológicas, facilitando la comprensión de los mecanismos alterados en la enfermedad de Parkinson.

3.8. *Integración con la red centrada en SNCA*

Una vez caracterizados los módulos de coexpresión y sus respectivas funciones biológicas, se procedió a integrar los resultados con una red centrada en el gen SNCA (α -sinucleína), ampliamente reconocido como uno de los principales actores moleculares implicados en la enfermedad de Parkinson.

El propósito de esta etapa fue evaluar la relación entre los módulos detectados y las interacciones moleculares que involucran a SNCA, con el fin de identificar posibles genes conectores o rutas compartidas que pudieran explicar mecanismos patológicos de la enfermedad.

3.8.1. *Obtención de la red de SNCA*

La red centrada en SNCA se generó utilizando la plataforma STRING (Search Tool for the Retrieval of Interacting Genes/Proteins), seleccionando el organismo *Homo sapiens* y un score de interacción mínimo de 0.7 (alta confianza).

Se estableció un límite de entre 50 y 150 nodos para capturar tanto las interacciones directas de SNCA como las conexiones de segundo orden (vecinos cercanos), manteniendo la red interpretable sin perder información relevante.

3.8.2. *Superposición con los módulos de coexpresión*

Una vez generada la red de SNCA, se compararon los genes pertenecientes a los módulos de coexpresión con los nodos de la red STRING.

Este cruce permitió identificar los genes de la red de coexpresión que interactúan directamente o a través de uno o dos intermediarios con SNCA, lo cual es especialmente relevante para priorizar candidatos con posible implicancia en la patogénesis del Parkinson.

3.9. *Visualización e integración final en Cytoscape*

Tras la obtención de los módulos de coexpresión y la red centrada en SNCA, se utilizó la plataforma Cytoscape para realizar la visualización e integración final de toda la red génica.

3.9.1. *Importación y estructura de la red*

Los archivos generados en Python (Coexpr_nodes.tsv) fueron importados a Cytoscape como nodos y aristas, respectivamente.

Cada nodo representa un gen, mientras que cada arista indica una correlación significativa entre dos genes ($|r| \geq 0.7$).

Posteriormente, se integró la información proveniente de STRING, incorporando los genes del entorno de SNCA y permitiendo unificar la red de coexpresión con las interacciones conocidas experimentalmente.

3.9.2. *Cálculo de medidas topológicas*

Dentro de Cytoscape se pueden calcular indicadores de centralidad, principalmente el grado (degree) y la intermediación (betweenness), con el fin de evaluar la importancia topológica de cada gen dentro de la red.

Los genes con valores altos en estas métricas fueron considerados nodos clave o hubs, ya que actúan como puntos de conexión o mediadores entre diferentes módulos.

Se priorizó el uso del grado de conectividad (degree) como criterio visual, al ser más intuitivo y estable en redes de tamaño medio.

De esta forma, los genes con mayor número de interacciones se destacaron automáticamente en el grafo, facilitando la identificación de posibles reguladores centrales.

3.9.3. *Interpretación y valor del enfoque visual*

La representación final en Cytoscape permitió observar de manera integrada la arquitectura de la red, mostrando cómo los módulos coexpresados se interrelacionan y en qué posición se encuentra SNCA dentro del sistema.

Gracias a esta visualización fue posible distinguir módulos densamente conectados, identificar genes puente entre comunidades y reconocer a SNCA como un nodo de alta conectividad, coherente con su papel central en la fisiopatología del Parkinson.

3.10. Priorización de genes candidatos

Una vez integradas las redes de coexpresión y de interacciones centradas en SNCA, se implementó una estrategia de priorización de genes candidatos con el objetivo de identificar aquellos genes que podrían desempeñar un papel relevante en la fisiopatología de la enfermedad de Parkinson.

Este paso buscó combinar la información estadística, topológica y biológica obtenida en las etapas anteriores en un único sistema de puntuación reproducible y transparente.

3.10.1. Criterios de priorización

Cada gen fue evaluado según tres criterios complementarios, que reflejan distintos aspectos de su relevancia dentro de la red y del contexto funcional:

- A. Pertenencia a un módulo enriquecido funcionalmente (+1): Se asignó un punto a los genes incluidos en módulos cuya función biológica resultó significativa (p -ajustada < 0.05) en los análisis de enriquecimiento.
- B. Alta conectividad o centralidad en la red (+1): Se otorgó un punto adicional a los genes con grado de conectividad alto para considerarse elementos estructuralmente relevantes que podrían regular o coordinar la expresión de otros genes en su módulo.
- C. Proximidad con SNCA en la red de interacciones (+1): Finalmente, se añadió un punto a los genes que se encontraban directamente conectados con SNCA o a una distancia máxima de tres nodos, según las interacciones detectadas en Cytoscape.

3.10.1. Asignación del puntaje total

El puntaje final de cada gen se obtuvo sumando los valores de los tres criterios anteriores, con un rango posible de 0 a 3 puntos:

Score	interpretación
3	Alta prioridad (presente en el módulo enriquecido, hub y cercano a SNCA)
2	Prioridad media (cumple dos criterios)
1	Evidencia limitada (cumple un criterio)
0	No cumple

Tabla 2. Puntaje total

3.11. De análisis bioinformático a propuestas biológicas

La última etapa metodológica consistió en traducir los hallazgos computacionales en hipótesis biológicas accionables, conectando los módulos, las rutas enriquecidas y los genes priorizados con posibles mecanismos moleculares implicados en la enfermedad de Parkinson.

El objetivo de esta fase fue proponer interpretaciones funcionales y potenciales líneas de investigación experimental basadas en la integración de los análisis realizados.

4. Resultados

4.1. Preparación y estructura final del dataset

Luego del proceso de normalización, filtrado y ordenamiento de los datos, con Galaxy y Python, obtenidos desde GEO, se conformó la matriz de expresión génica definitiva, utilizada como base para los análisis de reducción de dimensionalidad, co-expresión y enriquecimiento funcional.

Esta matriz final está compuesta por 26 columnas y 22.827 filas.

- Cada columna (excepto la primera) representa una muestra biológica individual, identificada por su código GSMxxxxx, correspondiente a tejido de sustancia negra humana, con condiciones de control y enfermedad de Parkinson (PD).
- La primera columna contiene el símbolo oficial del gen (por ejemplo, ATP5F1B, NDUFA12, RPL4, COX6C), que actúa como identificador único.
- Cada fila, por tanto, representa la expresión cuantitativa de un gen específico a lo largo de las 25 muestras analizadas.

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
Gene Symbol	GSM184354	GSM184355	GSM184356	GSM184357	GSM184358	
A1BG	21.31	25.12	36.87	36.16	47.03	
A1BG-AS1	12.05	5.57	5.65	6.65	66.74	
A1CF	13.474999999999998	15.955000000000002	14.41	24.06	28.665	
A2M	1563.82	564.74	447.38500000000005	831.13	806.5450000000001	
A2M-AS1	82.29	103.38	55.97	43.94	108.69	
A2ML1	16.09	14.28	8.95	26.255	31.765	
A2MP1	8.31	19.85	16.87	16.15	5.65	
A4GALT	34.28	6.58	10.98	10.53	22.2	
A4GNT	33.49	22.89	26.29	15.62	34.37	
AA06	18.94	26.81	23.28	41.04	93.93	
AAAS	49.94	95.48	52.33	4.32	33.7	
AACS	116.66	145.44	174.9	149.05	99.94	
AACSP1	7.58	16.18	0.41	2.06	20.42	
AADAC	1.32	13.3	2.03	3.59	35.77	
AADACL2	10.74	9.49	0.57	6.14	1.79	
AADACP1	0.47	1.2	1.9	0.46	12.17	
AADAT	79.18	109.85	68.27	49.61	110.03	
AAED1	125.38	90.48	129.99	138.29	97.88	
AAGAB	56.386666666666668	64.74333333333333	95.92666666666666	43.31333333333333	73.35000000000001	47.6761
AAK1	145.12714285714284	192.8857142857143	160.8857142857143	141.16	163.53428571428572	162.657

Ilustración 1. Dataset resumido

4.2. Análisis y distribución de las muestras usando PCA

En la Ilustración 2, se observa la distribución de las muestras correspondientes a sujetos control (azul) y con enfermedad de Parkinson (rojo) en función de las dos primeras componentes principales (PC1 y PC2). La PC1 explicó el 15.09 % de la variabilidad total y la PC2 el 8.52 %, lo que representa un 23.61 % de la varianza acumulada (ilustración 3).

Si bien las muestras de ambos grupos presentan cierta superposición, se aprecia una tendencia a la separación parcial entre los individuos control y los pacientes con Parkinson, especialmente a lo largo del eje de la PC1.

Esto sugiere la presencia de patrones de expresión diferencial que podrían estar asociados al estado patológico, aunque todavía no de manera completamente definida, probablemente debido a la heterogeneidad intrínseca del Parkinson y a la variabilidad biológica entre los sujetos.

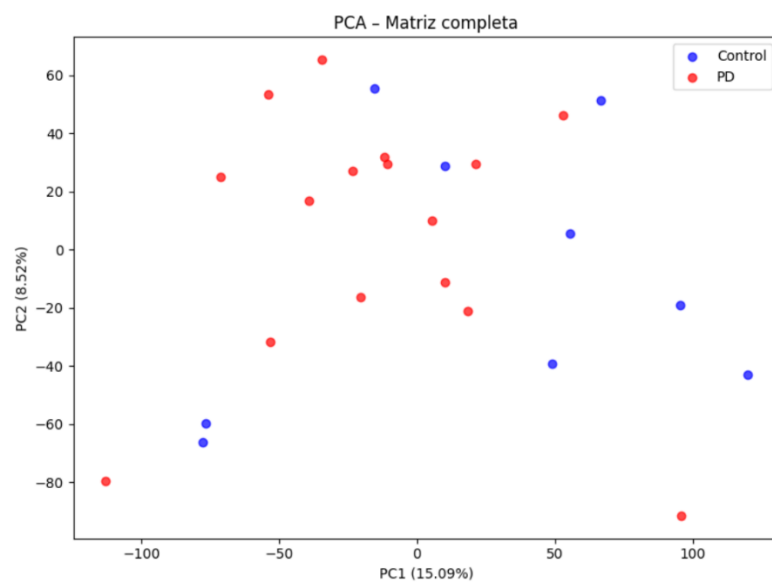


Ilustración 2. PCA

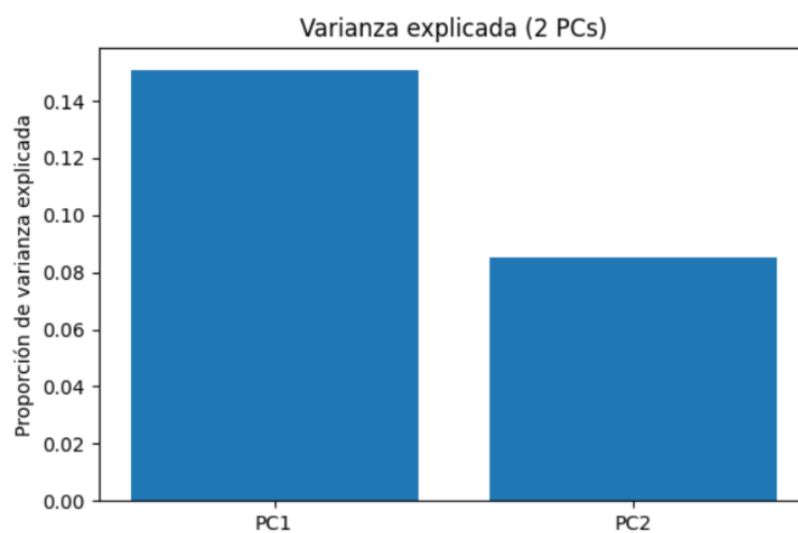


Ilustración 3. Varianza PCA

4.3. *K – means*

En primer lugar, se empleó el algoritmo K-Means con $k = 2$, correspondiente al número esperado de grupos biológicos (Control y PD).

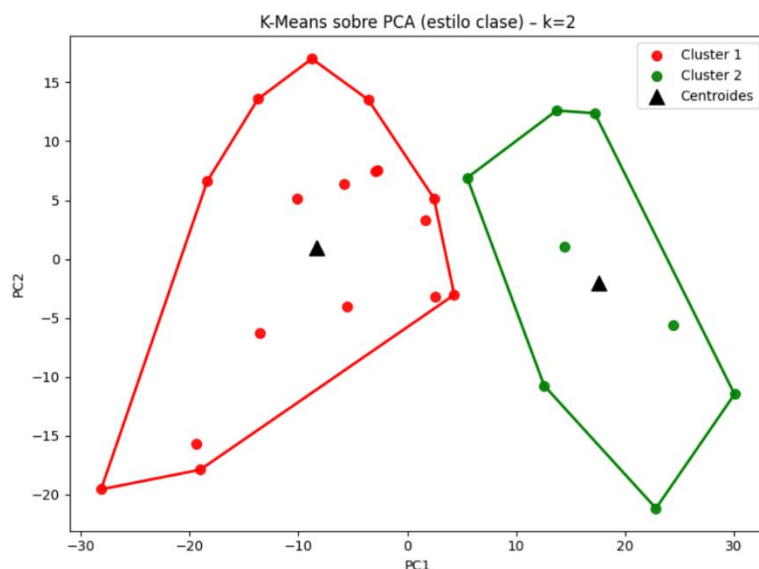


Ilustración 4. Clustering

En la Ilustración 4, se observan los dos clusters obtenidos proyectados sobre el espacio PCA. Cada grupo se muestra con un color distinto (rojo y verde), y los triángulos negros representan los centroides de cada clúster.

El modelo logró una separación visible entre ambos grupos, consistente con la tendencia observada en el análisis de componentes principales. Esto indica que las muestras comparten patrones de expresión diferenciados, compatibles con el estado patológico o control.

Posteriormente, se aplicó un análisis de clustering jerárquico, utilizando tres métodos de enlace (single, complete y average linkage) para verificar la estabilidad de la clasificación.

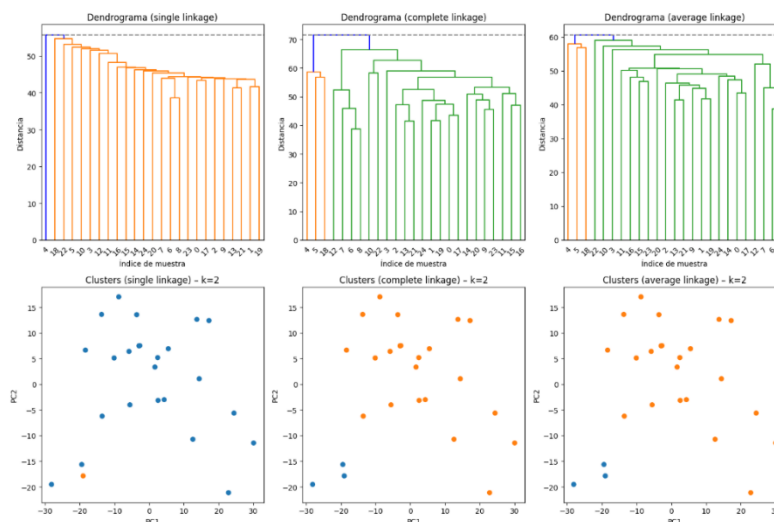


Ilustración 5. Dendogramas

Los resultados se muestran en la Ilustración 5, donde los dendrogramas permiten observar cómo las muestras tienden a agruparse en dos grandes ramas, manteniendo una organización coherente con los grupos obtenidos por K-Means.

Entre los métodos evaluados, el average linkage mostró una estructura más equilibrada y biológicamente interpretable, por lo que fue seleccionado para los análisis posteriores.

Finalmente, en la Ilustración 6 se presenta un mapa de calor global con los 1.500 genes más variables, estandarizados por fila (z-score).

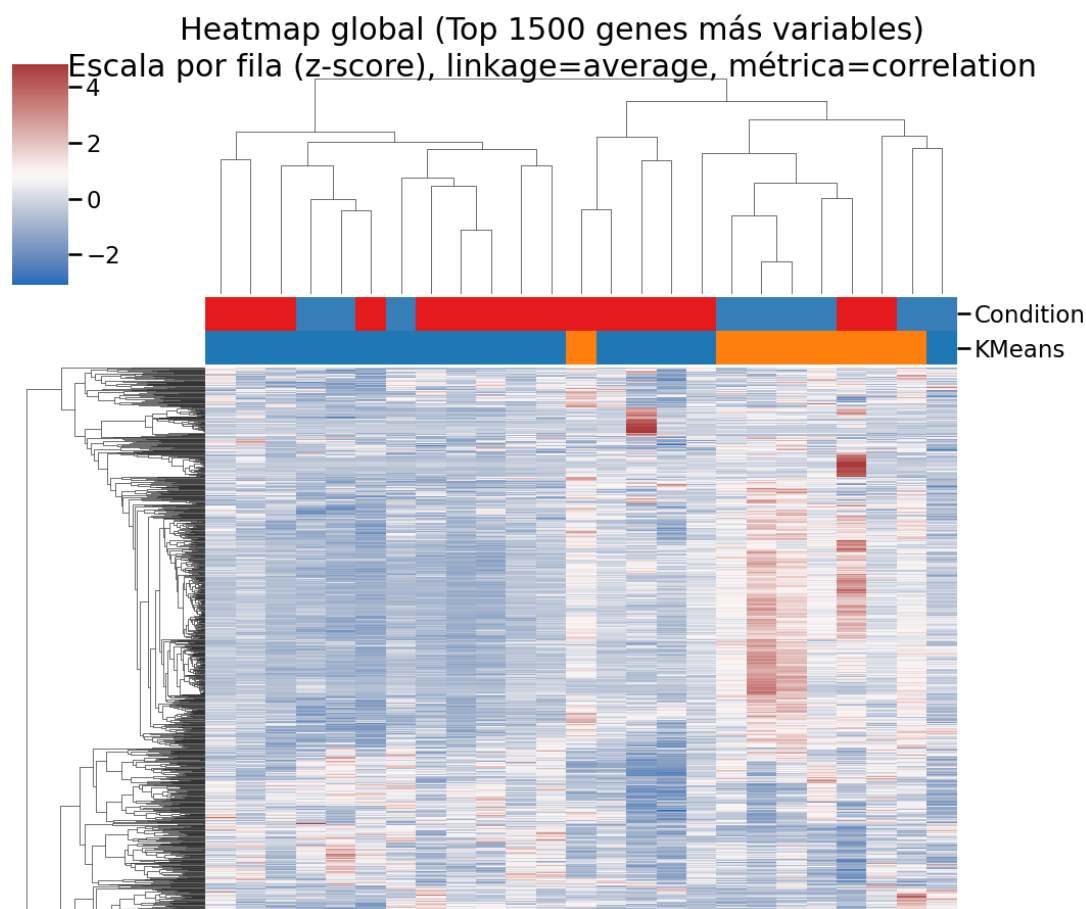


Ilustración 6. Heatmap Global de 1500 genes

Los colores indican la intensidad relativa de expresión (rojo: sobreexpresión; azul: subexpresión), mientras que las barras superiores representan las etiquetas de condición y los clusters asignados por K-Means.

4.4. Selección de genes informativos y red de coexpresión

Para reducir la dimensionalidad y concentrar el análisis en los genes más relevantes, se realizó un filtrado por varianza, seleccionando los 500 genes con mayor variabilidad de expresión entre las muestras. Este paso permite descartar genes con escasa fluctuación (que aportan poco

poder discriminante) y mantener únicamente aquellos que reflejan cambios biológicamente significativos entre los grupos control y Parkinson.

A continuación, se generó un mapa de calor (Ilustración 7) con estos 500 genes, escalados por fila mediante z-score y ordenados según un clustering jerárquico promedio. Los colores representan niveles relativos de expresión (rojo = sobreexpresión, azul = subexpresión).

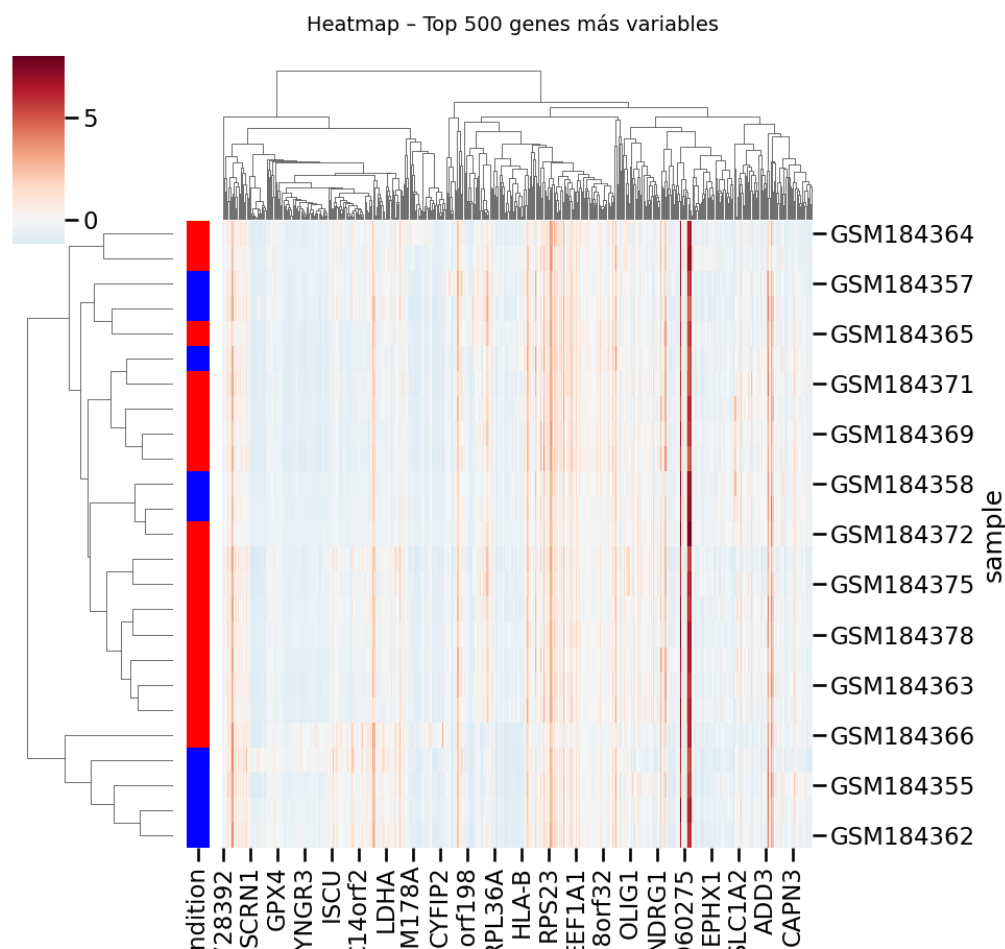


Ilustración 7. Heatmap con 500 genes

En el mapa se observan patrones de expresión diferenciados entre los individuos con enfermedad de Parkinson (en rojo en la barra lateral) y los controles (en azul), lo que sugiere la existencia de subconjuntos de genes regulados de forma coordinada según la condición biológica.

Posteriormente, para cuantificar las relaciones entre los genes y detectar posibles módulos de coexpresión, se calculó la correlación de Pearson entre todos los pares de genes seleccionados. Este coeficiente mide la fuerza y dirección de la asociación lineal entre dos variables de expresión:

- Un valor r cercano a +1 indica que ambos genes aumentan o disminuyen conjuntamente,

- Un valor r cercano a -1 sugiere que la expresión de uno aumenta cuando el otro disminuye,
- Y valores r próximos a 0 reflejan ausencia de relación lineal.

Se conservaron únicamente las correlaciones con $|r| \geq 0.85$ y significativas ($p < 0.05$), construyendo así una matriz de adyacencia que sirvió de base para la red de coexpresión génica (Ilustración 8).

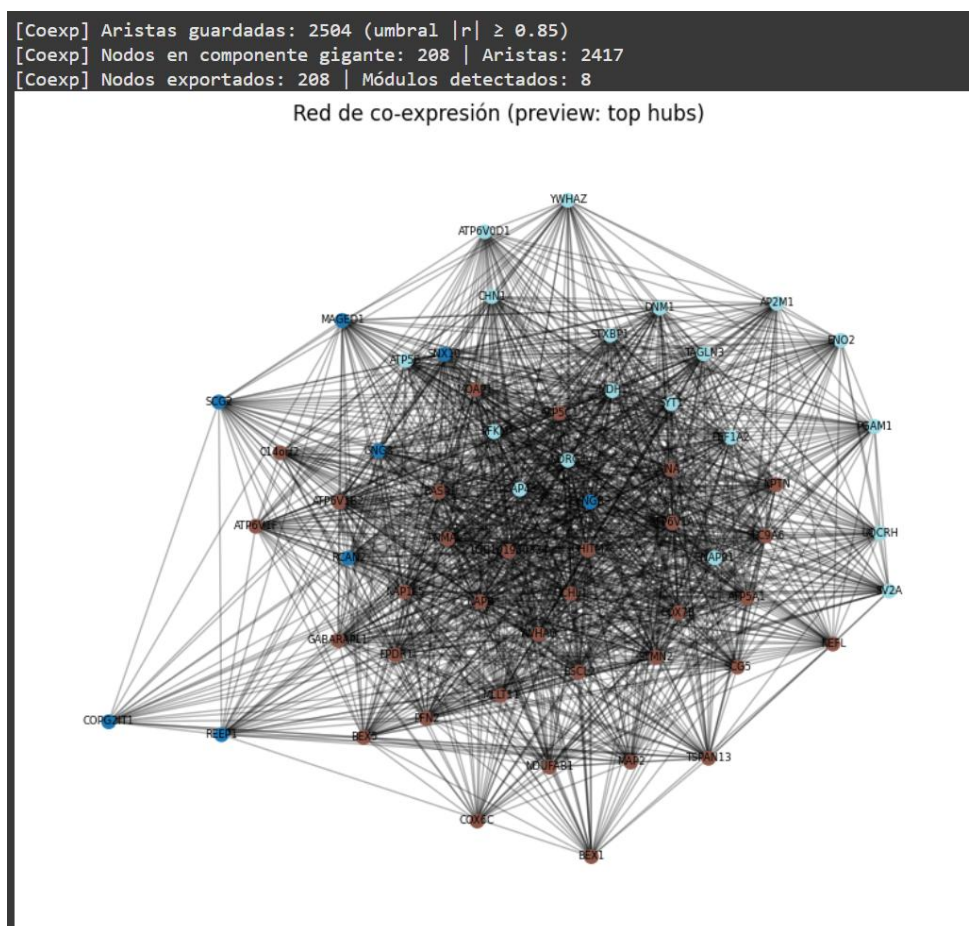


Ilustración 8. Red de Co - expresión

En dicha red, cada nodo representa un gen, y cada arista (conexión) representa una correlación fuerte entre dos genes.

El análisis reveló un componente gigante formado por 208 genes interconectados mediante 2.417 aristas, organizados en 8 módulos principales detectados con el algoritmo de Louvain. Estos módulos agrupan genes que tienden a expresarse conjuntamente, lo que sugiere participación en rutas o procesos biológicos compartidos.

4.5. *Análisis funcional de los módulos de coexpresión*

Cada módulo representa un conjunto de genes con patrones de expresión similares, presumiblemente involucrados en procesos biológicos comunes.

El análisis de enriquecimiento funcional en STRING permitió asignar a cada módulo un perfil biológico distintivo, aportando una visión integradora de los mecanismos potencialmente alterados en la enfermedad de Parkinson (EP).

Cada módulo fue analizado de manera independiente, aplicando enriquecimiento funcional sobre los conjuntos de genes coexpresados para identificar procesos biológicos (GO: Biological Process), rutas metabólicas (KEGG) y componentes celulares con sobre-representación significativa ($FDR < 0.05$).

A continuación, se detallan los principales hallazgos para los módulos más relevantes.

4.5.1. *Modulo M1: Núcleo funcional energético*

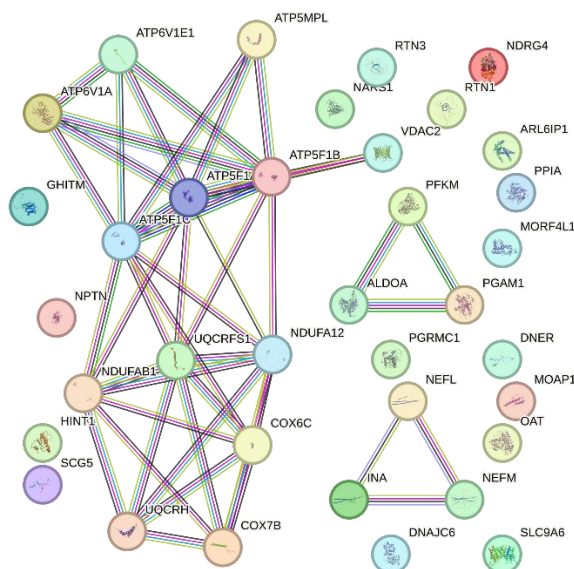


Ilustración 9. M1

El módulo M1 presentó un enriquecimiento altamente significativo en procesos biológicos relacionados con la producción de energía y la función mitocondrial, destacándose los términos “Oxidative phosphorylation”, “ATP biosynthetic process”, “Mitochondrial ATP synthesis coupled electron transport” y “Aerobic electron transport chain” ($FDR < 10^{-7}$).

Los genes predominantes pertenecen a subunidades de los complejos mitocondriales I, III, IV y V (NDUFA12, UQCRC1, COX6C, ATP5F1B, ATP6V1A), junto con proteínas accesorias que participan en el mantenimiento del potencial mitocondrial y el transporte de electrones (VDAC2, GHITM, COX7B).

Estos resultados reflejan una actividad coordinada de la cadena respiratoria y la síntesis de ATP, procesos críticos para la homeostasis neuronal. En el contexto de la enfermedad de Parkinson, las alteraciones en estas rutas se asocian con disfunción mitocondrial, estrés oxidativo y pérdida de neuronas dopaminérgicas.

Por tanto, el M1 podría representar un núcleo funcional bioenergético, clave para el mantenimiento del metabolismo oxidativo neuronal y potencialmente sensible a la disfunción mitocondrial inducida por α -sinucleína (SNCA).

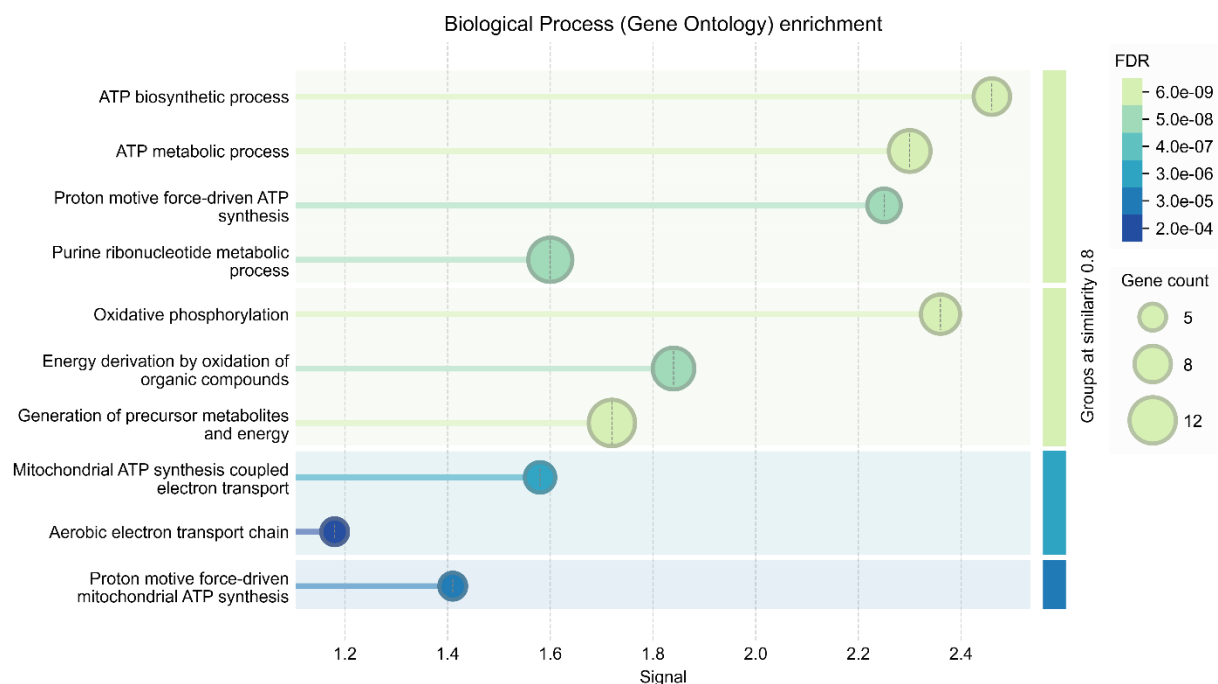


Ilustración 10. Procesos Biológicos M1

4.5.2. M2: Núcleo funcional de traducción y de actividades ribosomales

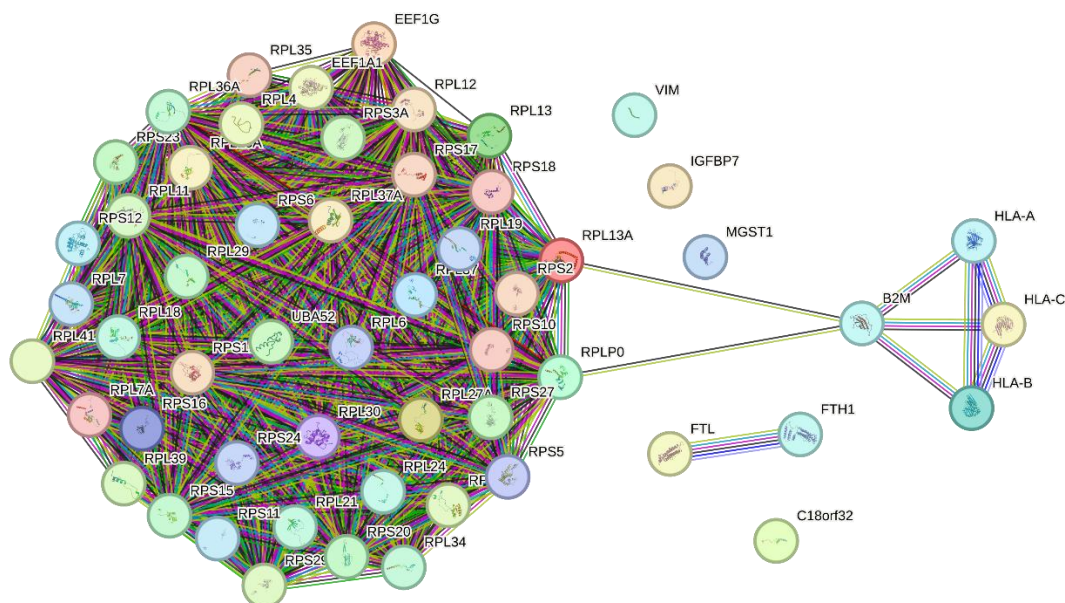


Ilustración 11. M2

El módulo M2 presentó un enriquecimiento altamente significativo en procesos vinculados a la traducción proteica y la biogénesis ribosomal, destacándose los términos “Cytoplasmic translation”, “Ribosome assembly”, “Ribosome biogenesis” y “rRNA processing” ($FDR < 10^{-77}$).

Estos resultados sugieren que M2 representa un núcleo de síntesis proteica activa, esencial para el mantenimiento de la homeostasis neuronal. En el contexto de la enfermedad de Parkinson, las alteraciones en la maquinaria de traducción y en la función ribosomal se asocian con estrés proteostático y acumulación de α -sinucleína (SNCA), contribuyendo al daño celular progresivo.

Por tanto, el Módulo M2 podría vincularse con la regulación de la traducción y la respuesta al estrés del retículo endoplásmico, constituyendo un clúster clave para comprender los mecanismos de alteración proteica observados en la neurodegeneración.

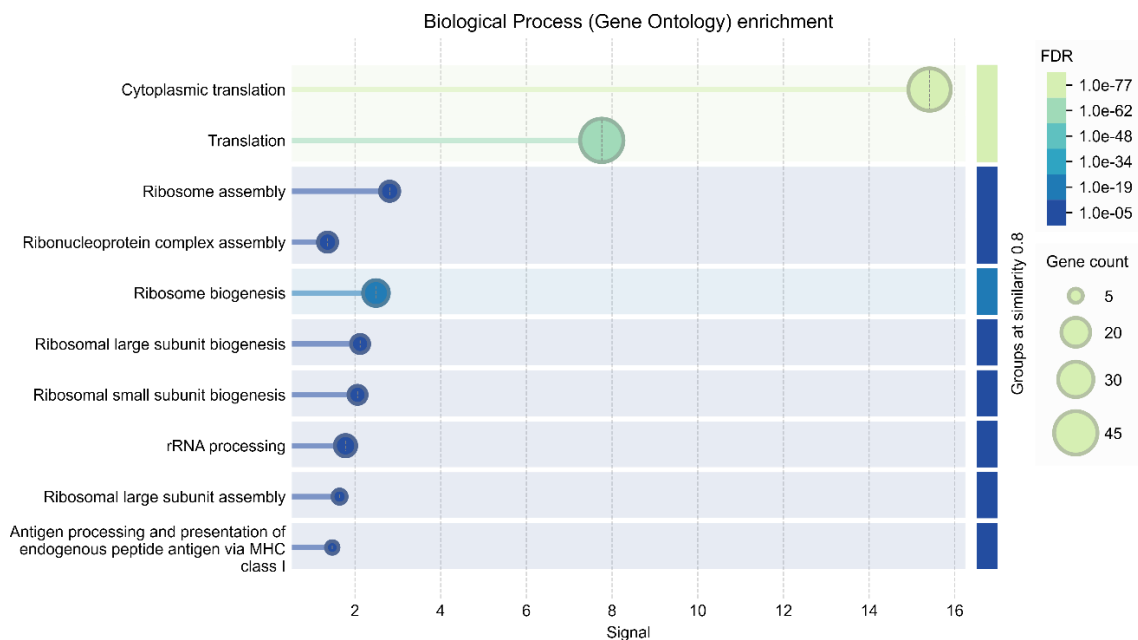


Ilustración 12. Procesos Biológicos M2

4.5.3. M3: Núcleo funcional de dinámica del citoesqueleto

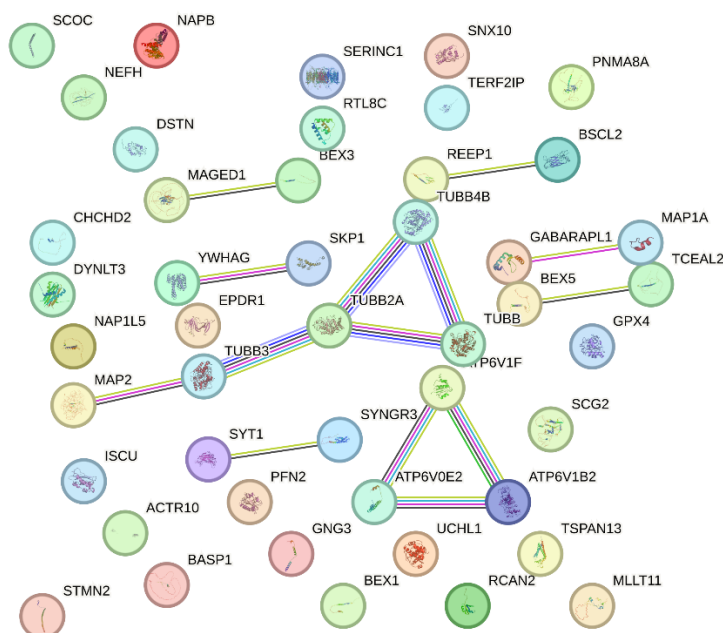


Ilustración 13. M3

El Módulo M3 presentó un enriquecimiento biológico específico en el proceso “*Microtubule-based process*” ($FDR < 0.01$), involucrando genes asociados a la organización y estabilidad del citoesqueleto neuronal.

En la red de co-expresión (Ilustración 13), se observan interacciones densas entre genes de las familias de tubulinas (TUBB2A, TUBB3, TUBB4B) y proteínas asociadas a microtúbulos, como *MAP1A*, *SYT1*, *GABARAPL1*, *PFN2* y *DYNLT3*. Estas proteínas cumplen funciones esenciales en el transporte axonal, la formación de sinapsis y el mantenimiento estructural de las neuronas dopaminérgicas.

El predominio de genes relacionados con la dinámica microtubular sugiere que M3 representa un módulo funcional neuronal, íntimamente ligado a la integridad del citoesqueleto. En el contexto de la enfermedad de Parkinson, la disfunción en los microtúbulos y en el transporte axonal se ha asociado con la degeneración progresiva de neuronas dopaminérgicas y con la alteración en la distribución de proteínas como α -sinucleína y *parkin*.

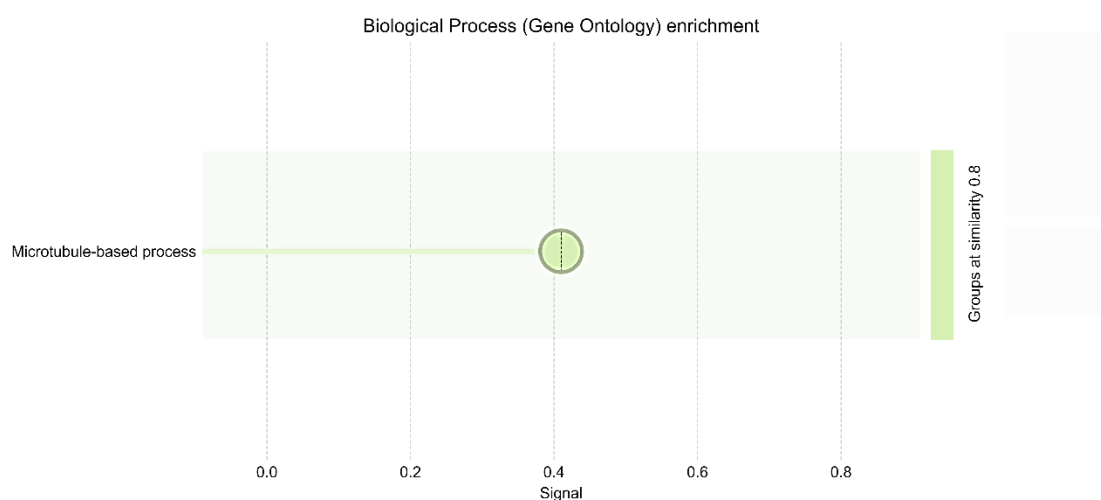


Ilustración 14. Procesos Biológicos M3

4.5.4. M4: Núcleo funcional mitocondrial respiratorio

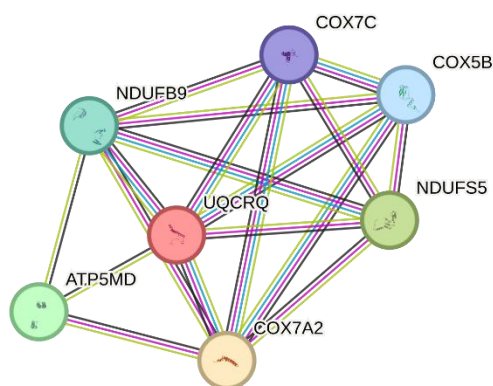


Ilustración 15. M4

El Módulo M4 exhibe un enriquecimiento biológico altamente significativo en procesos relacionados con la cadena de transporte de electrones y la fosforilación oxidativa ($FDR < 10^{-6}$). Entre los términos destacados se incluyen “Aerobic electron transport chain”, “Mitochondrial ATP synthesis coupled electron transport” y “Proton motive force-driven ATP synthesis”.

La red de interacción (Ilustración 15) está conformada por genes codificantes de componentes de los complejos mitocondriales I, III, IV y V, tales como NDUFB9, NDUFS5, UQCRCQ, COX5B, COX7A2, COX7C y ATP5MD. Estas proteínas participan en el flujo de electrones y la generación del gradiente electroquímico de protones necesario para la síntesis de ATP, siendo esenciales para el metabolismo energético neuronal.

Desde un punto de vista fisiopatológico, el mal funcionamiento de los complejos respiratorios mitocondriales constituye uno de los mecanismos más ampliamente documentados en la enfermedad de Parkinson. La disrupción en la cadena de transporte de electrones incrementa la producción de especies reactivas de oxígeno (ROS) y provoca estrés oxidativo, afectando directamente la supervivencia de las neuronas dopaminérgicas.

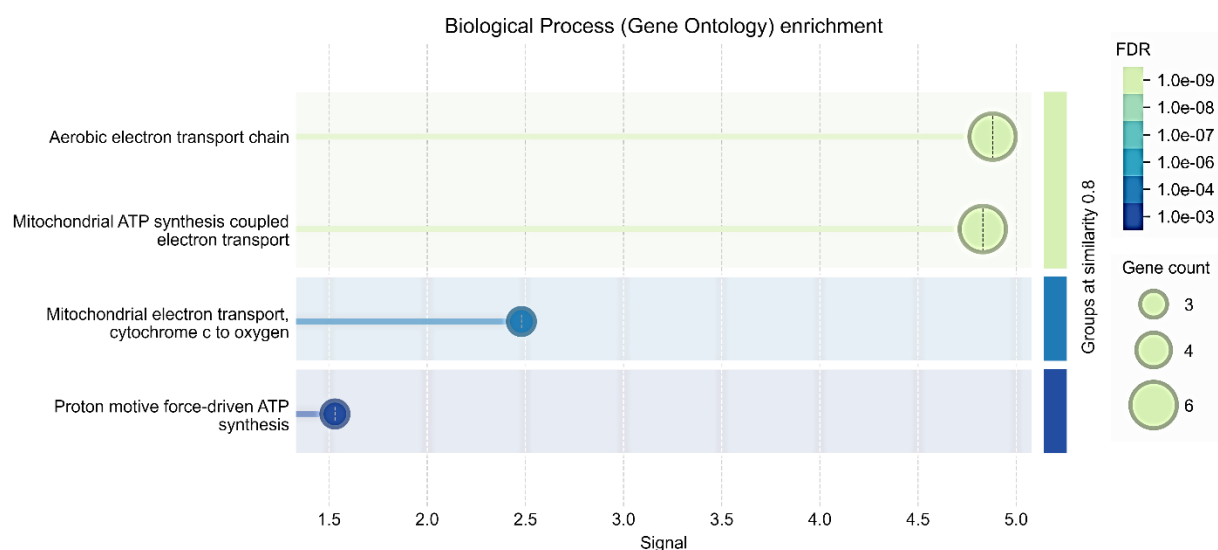


Ilustración 16. Procesos Biológicos M4

4.5.5. M5: Núcleo funcional sináptico y de señalización neuronal

El Módulo M5 mostró un enriquecimiento significativo en procesos asociados a la transmisión sináptica y al ciclo de vesículas sinápticas ($FDR < 10^{-4}$). Entre los términos más relevantes se destacan “Synaptic vesicle cycle”, “Synaptic vesicle endocytosis”, “Chemical synaptic transmission” y “Nervous system development”.

La red de interacción (Ilustración 17) está compuesta principalmente por genes involucrados en la liberación y reciclaje de neurotransmisores, tales como *SNAP25*, *DNM1*, *STXBPI*, *SYNPR*,

GABRA1, *GABBR2* y *SLC17A6*. Estas proteínas participan en la exocitosis y endocitosis vesicular, la señalización gabaérgica y glutamatergica, y la regulación del flujo de iones a través de las membranas neuronales.

Este módulo agrupa genes que sostienen funciones críticas en la plasticidad sináptica, la coordinación del tráfico vesicular y la homeostasis del calcio, elementos fundamentales para la comunicación neuronal eficiente. Alteraciones en estas vías se han vinculado con la disminución de la liberación dopaminérgica

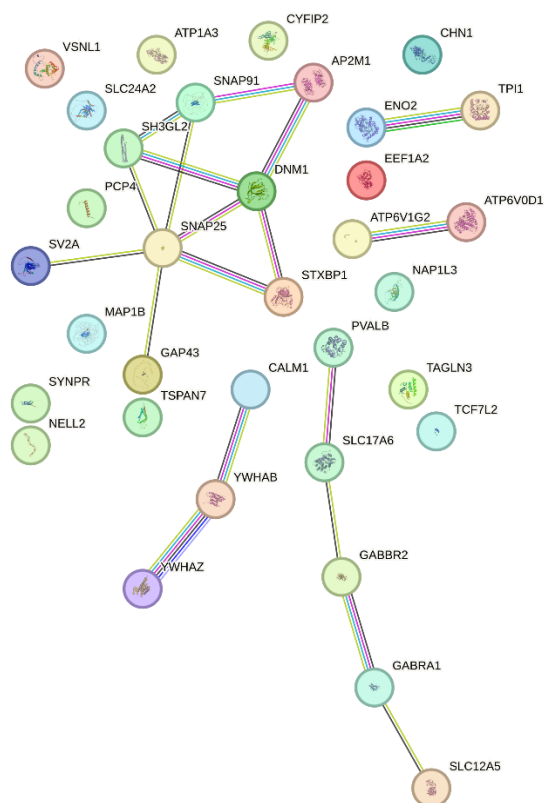


Ilustración 17. M5

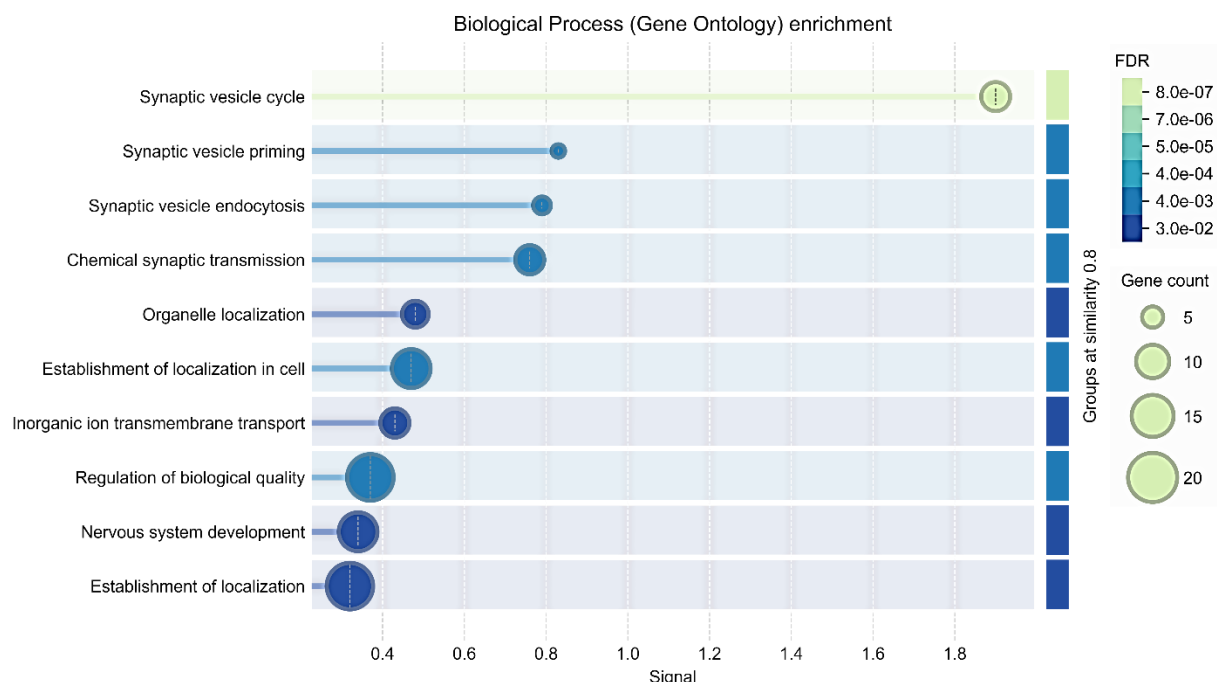


Ilustración 18. Procesos Biológicos M5

4.5.6. M6: transporte y metabolismo de neurotransmisores

El Módulo M6 mostró un enriquecimiento funcional fuertemente asociado al transporte y metabolismo de neurotransmisores, con términos como “*Neurotransmitter transport*”, “*Aminergic neurotransmitter loading into synaptic vesicle*” y “*Dopamine biosynthetic process*” ($FDR < 10^{-3}$).

La red de interacción (Ilustración 19) incluye genes característicos del metabolismo dopaminérgico, como TH (tirosina hidroxilasa), SLC6A3 (transportador de dopamina), SLC18A2 (transportador vesicular de monoaminas, VMAT2) y SCRN1. Estos genes conforman un eje funcional esencial para la síntesis, empaquetamiento y liberación de dopamina en las neuronas presinápticas.

El enriquecimiento observado indica que este módulo agrupa genes cuya coexpresión refleja la actividad específica de las neuronas dopaminérgicas, las más afectadas en la enfermedad de Parkinson. La disfunción de estos genes puede traducirse en reducción de la señalización dopaminérgica, estrés oxidativo y acumulación de dopamina citosólica, fenómenos clave en la neurodegeneración del sistema nigroestriatal.

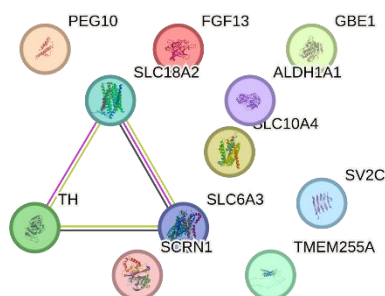


Ilustración 19. M6

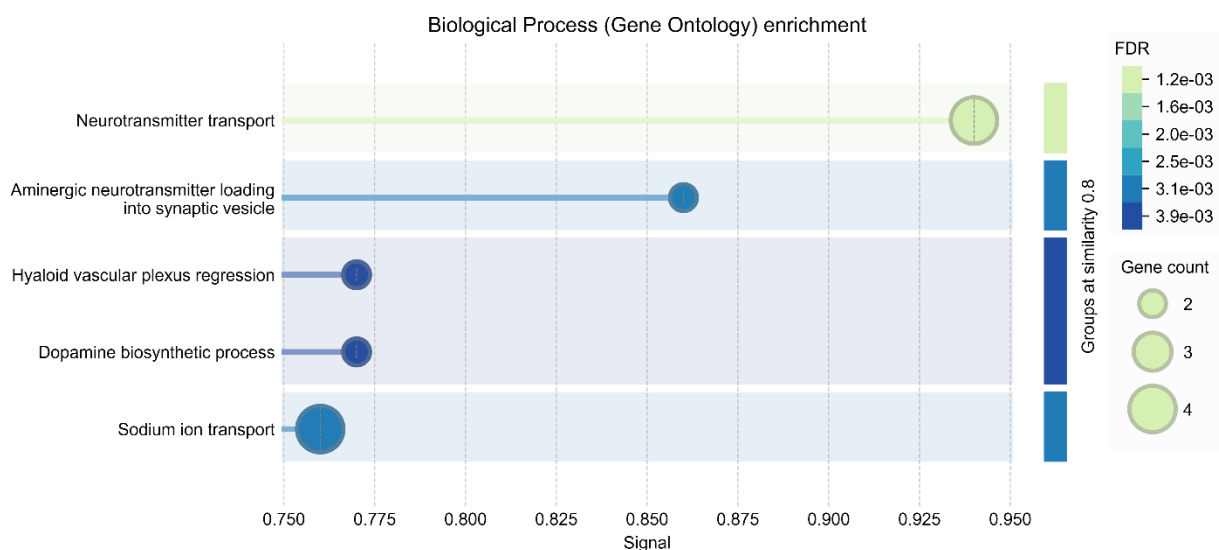


Ilustración 20. Procesos Biológicos M6

4.5.7. *M7: submodulo sináptico con evidencia limitada*

Este módulo pequeño (≈ 6 genes) mostró un enriquecimiento único asociado a localización a gránulo secretorio/Chromogranins (señal ~ 1.0). Los genes CHGB, SYT4, HPCAL1 sugieren un componente granular-sináptico coherente con M5 – M6. Sin embargo, el bajo tamaño y la conectividad reducen su robustez. Lo considero un submódulo consistente pero secundario, y no altera las conclusiones principales.

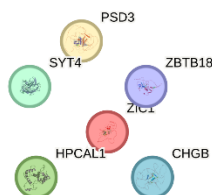


Ilustración 21. M7

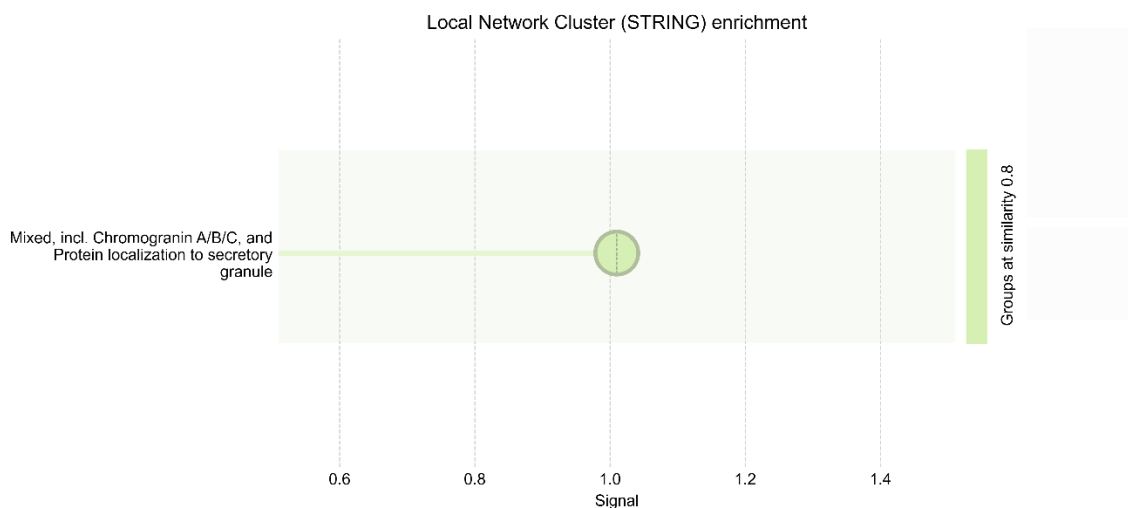


Ilustración 22. Proceso Biológico M7

4.5.8. M8 y M9

M8 y M9 no mostraron enriquecimiento significativo ni conectividad robusta, por lo tanto, se omiten por criterios de calidad para evaluar en STRING.

4.6. Resultados finales: integración global de la red de coexpresión

4.6.1. Integración funcional de los módulos M1–M6

La integración de los módulos permitió construir una red funcional global (Ilustración 23) donde convergen los procesos más representativos de la enfermedad de Parkinson. La red revela una arquitectura jerárquica bien definida, con tres niveles funcionales interconectados:

- A. Nivel central – Módulos M1 y M2 (Proteostasis y traducción): agrupan genes ribosomales y de biosíntesis proteica (translation, peptide biosynthetic process), conformando el núcleo estructural de la red.
- B. Nivel intermedio – Módulos M3 y M4 (Metabolismo energético): concentran subunidades mitocondriales y procesos de fosforilación oxidativa (oxidative phosphorylation, ATP metabolic process), actuando como sistema de soporte metabólico.
- C. Nivel periférico – Módulos M5 y M6 (Función sináptica y dopaminérgica): integran genes vinculados a la transmisión neuronal (synaptic vesicle cycle, dopamine biosynthetic process), representando la salida funcional de la red hacia la actividad neuronal.

El análisis de enriquecimiento combinado (Ilustración 24) confirma esta jerarquía: los módulos se agrupan en torno a tres ejes biológicos, como proteostasis, metabolismo energético y neurotransmisión, que se interrelacionan de manera estrecha.

Esta organización sugiere que las alteraciones en la síntesis proteica y la bioenergética mitocondrial pueden repercutir directamente sobre la homeostasis sináptica dopaminérgica, reflejando la interdependencia molecular característica del proceso neurodegenerativo asociado a SNCA.

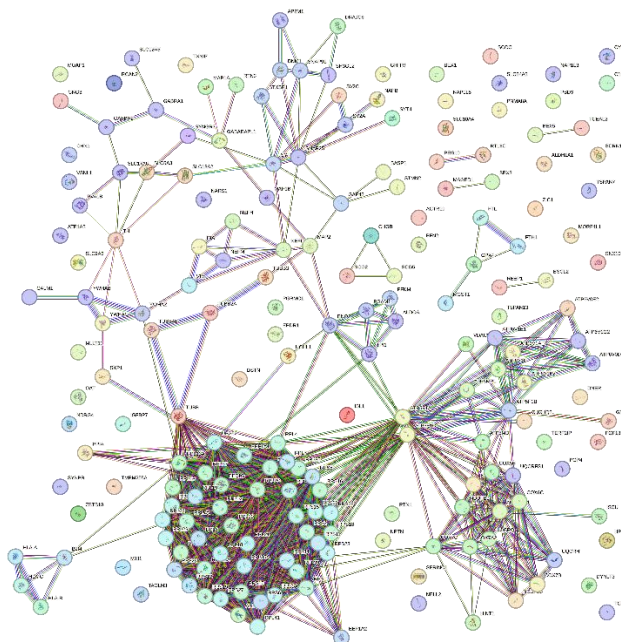


Ilustración 23. Red Integrada M1-M6

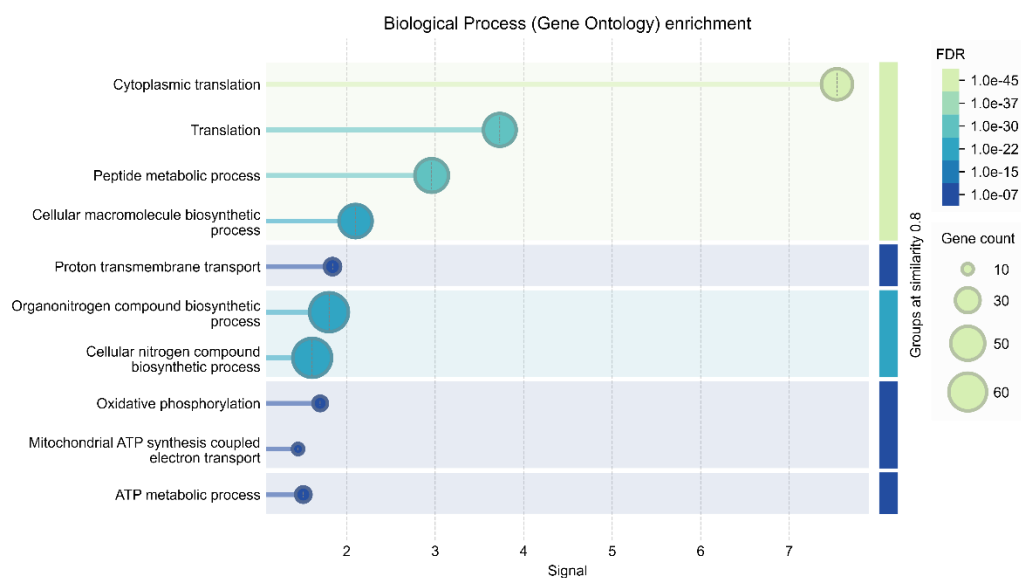
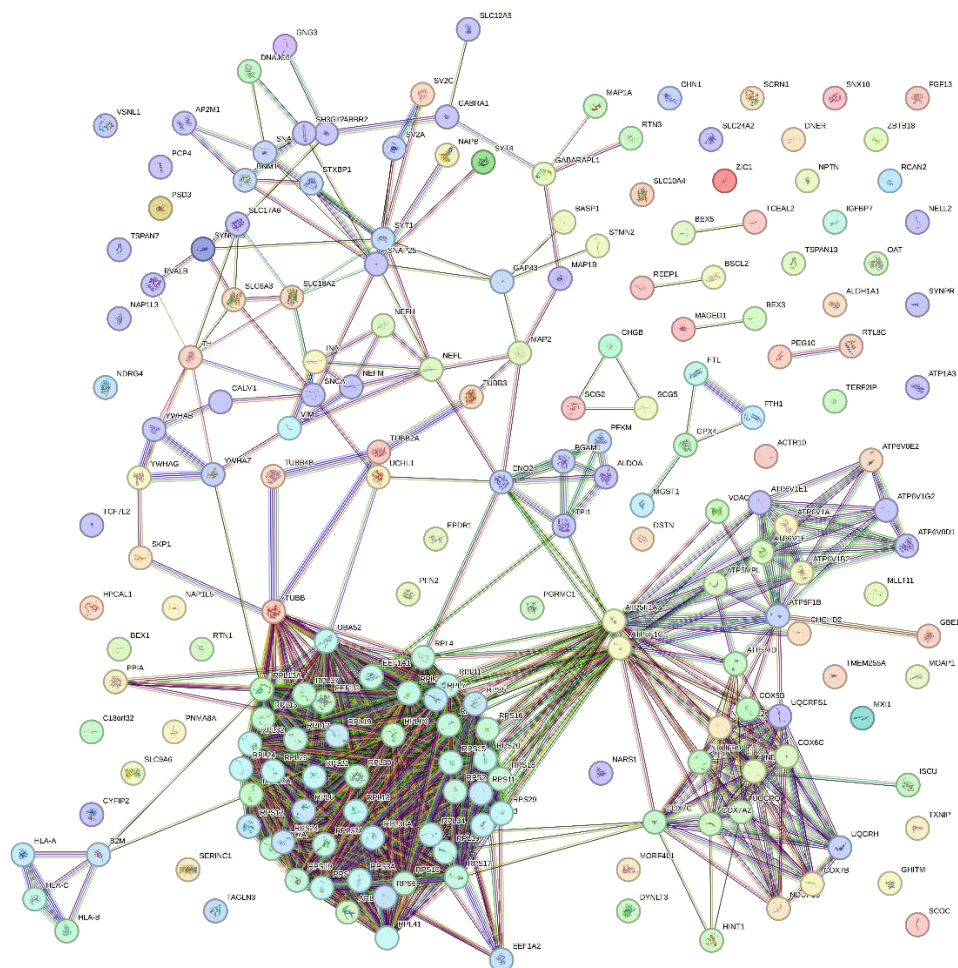


Ilustración 24. Procesos Biológicos generales

4.6.2. Integración M1-M6 con SNCA en STRING



4.6.3. Integración centrada en SNCA en Cytoscape

La integración final se realizó en Cytoscape, unificando los seis módulos funcionales (M1–M6) con la red centrada en SNCA, proteína clave en la patogénesis de la enfermedad de Parkinson, mediante la herramienta *Advanced Network Merge* utilizando la opción UNION.

4.6.4. Reasignación de módulos en la red integrada

Durante la integración de la red de coexpresión con la red centrada en SNCA, se produjo una reasignación automática de las etiquetas modulares. Este efecto se debe a que, al combinar ambas redes mediante la opción UNION en Cytoscape, los nodos importados desde STRING no poseían el atributo *Module*, lo que generó una propagación y reindexación alfabética de los valores existentes en esa columna.

En consecuencia, los rótulos modulares visibles en la red integrada no corresponden a los módulos originales de coexpresión, ya que las etiquetas fueron alteradas durante el proceso de fusión.

4.6.5. Identificación de hubs

Una vez integrada la red de coexpresión con las interacciones centradas en SNCA, se calcularon las métricas topológicas correspondientes al grado de conectividad (Degree) para cada nodo dentro de la red combinada.

Esta medida cuantifica el número de conexiones directas que posee un gen con otros elementos del sistema y se considera un indicador intuitivo de su relevancia estructural y funcional dentro de la red.

Los nodos con mayor grado fueron definidos como hubs estructurales, dado que concentran la mayor densidad de enlaces y contribuyen de manera determinante a la estabilidad global de la red.

Los diez genes con mayor grado de conectividad se presentan en la siguiente tabla:

Gen	Degree	Función biológica principal
NAPB	74	Proteína asociada al complejo SNARE; participa en la liberación sináptica de neurotransmisores.
COX7B	72	Subunidad del complejo IV mitocondrial; interviene en la respiración oxidativa y la producción de energía.
UCHL1	70	Enzima desubiquitinasa neuronal; regula la degradación de proteínas y la homeostasis sináptica
ATP6V1A	69	Subunidad A de la ATPasa V-tipo; controla el pH vesicular y la neurotransmisión.
PFKM	67	Fosfofructoquinasa muscular; cataliza una etapa clave de la glucólisis, reflejando la alta demanda energética neuronal.
EEF1A2	67	Factor de elongación 1- α 2; implicado en la síntesis proteica y la estabilidad del citoesqueleto.
INA	67	Neurofilamento α -internexina; componente estructural del citoesqueleto axonal.
YWHAG	65	Proteína adaptadora 14-3-3 γ ; modula la señalización intracelular y el transporte de proteínas fosforiladas.

BSCL2	63	Proteína seipina; participa en la organización de membranas y el mantenimiento de vesículas neuronales.
STXBP1	63	Proteína fijadora de syntaxina 1; esencial para la fusión de vesículas sinápticas.

Tabla 3. Tabla de hubs identificados

Estos genes definen un núcleo funcional neuronal y mitocondrial que integra procesos de transmisión sináptica, metabolismo energético, degradación proteica y control vesicular. Su alta conectividad indica que desempeñan un papel central en la red molecular analizada, y su alteración podría generar un impacto sistémico sobre múltiples rutas neuronales interdependientes.

Entre ellos, SNCA se posiciona como nodo integrador dentro de la estructura, estableciendo conexiones con varios de estos hubs y consolidando su papel como elemento clave en la conectividad global y la homeostasis neuronal.

4.7. Priorización de genes candidatos

Siguiendo con los criterios de priorización (3.10):

Gen	Modulo	Degree	Distancia a SNCA (nodos)	Puntaje total
NAPB	M3	74	8	2
COX7B	M1	72	6	2
UCHL1	M3	70	6	2
ATP6V1A	M1	69	5	2
PFKM	M1	67	8	2
EEF1A2	M5	67	2	3
INA	M1	67	4	2
YWHAG	M3	65	5	2
BSCL2	M3	63	4	2
STXBP1	M5	63	2	3

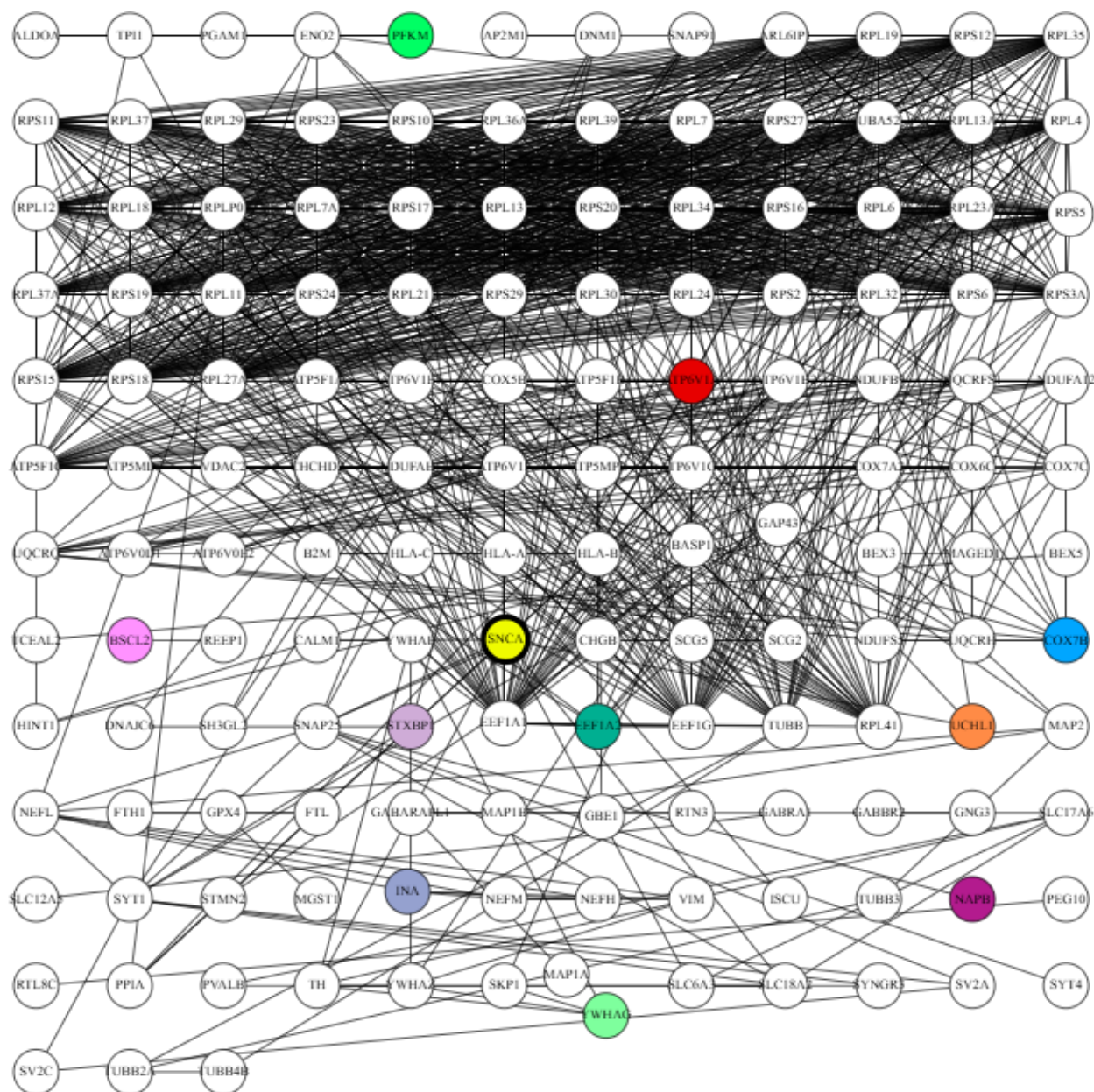


Ilustración 26. Priorización de genes candidatos

Los resultados muestran que los genes con mayor prioridad (EEF1A2 y STXBP1) pertenecen a módulos funcionales vinculados con la síntesis proteica y la neurotransmisión sináptica (M5), ambos procesos íntimamente relacionados con la fisiopatología del Parkinson.

Los demás hubs, aunque más alejados topológicamente de SNCA, se distribuyen en módulos mitocondriales (M1) y energéticos (M3), reflejando la convergencia entre metabolismo oxidativo y función neuronal. En conjunto, la priorización evidencia un núcleo de genes clave en la regulación del tráfico vesicular, el metabolismo energético y la homeostasis proteica, confirmando que la disfunción simultánea de estos mecanismos podría contribuir al desequilibrio celular que caracteriza a la neurodegeneración mediada por α -sinucleína.

4.8. *De análisis bioinformático a propuestas biológicas*

Como primera propuesta, habría que reorientar el estudio del Parkinson hacia una biología de redes celulares dinámicas, en la que la disfunción no reside únicamente en la acumulación de α -sinucleína, sino en la pérdida de sincronía entre módulos funcionales críticos. Desde esta perspectiva, la enfermedad podría entenderse como un desacople progresivo entre metabolismo energético y plasticidad sináptica.

La investigación futura debería centrarse en reconstruir esta sincronía mediante la co-modulación experimental de rutas mitocondriales y sinápticas, utilizando modelos neuronales que permitan medir no solo la viabilidad celular, sino también la restauración de la conectividad funcional entre ambas redes.

En la segunda propuesta busco que actuemos antes de que el daño ocurra, fortaleciendo los mecanismos naturales que tiene la célula para mantener sus proteínas en equilibrio. En lugar de intentar eliminar los agregados de α -sinucleína una vez que ya se formaron, la idea es entrenar a la célula para que tolere mejor el estrés proteico y responda de manera más eficiente. Esto implicaría estimular, de forma controlada, los procesos que reciclan o reparan proteínas, ayudando a prevenir el mal plegamiento desde etapas tempranas.

En otras palabras, se propone pasar de una visión de “corregir el daño” a una de “preparar al sistema para resistirlo”, entendiendo la neuroprotección como una forma de adaptación celular activa más que como una simple reparación posterior.

5. Conclusión

Este trabajo integró distintas herramientas de análisis bioinformático, topológico y funcional para comprender cómo SNCA, la proteína α -sinucleína, se inserta en la compleja red molecular de la enfermedad de Parkinson. A través del estudio de coexpresión génica y su integración con datos de interacción proteica, fue posible reconstruir 9 módulos de coexpresión que refleja la interdependencia entre procesos sinápticos, mitocondriales y de síntesis proteica, pilares del equilibrio neuronal.

La identificación de genes hubs (NAPB, UCHL1, ATP6V1A, STXBP1, etc.) altamente conectados y cercanos a SNCA reveló un conjunto de componentes que podrían influir de manera indirecta en su agregación o toxicidad, sugiriendo que el daño neuronal surge más de la ruptura de un sistema integrado que de la falla de un solo gen.

En conjunto, los resultados refuerzan la idea de que el Parkinson no puede explicarse por un único evento molecular, sino por la descoordinación progresiva de redes funcionales interdependientes que mantienen la homeostasis neuronal.

El enfoque adoptado, que combina la modelización computacional con la interpretación biológica, muestra el valor del análisis de redes como herramienta para transformar datos en hipótesis experimentales, acercando la bioinformática a la biología real de la enfermedad.

6. Bibliografía

National Center for Biotechnology Information. (2007). *GSE7621: Expression data of substantia nigra from postmortem human brain of Parkinson's disease patients (PD)*. GEO

Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J et al. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet* 2007 Jun;3(6):e98. PMID: [17571925](https://pubmed.ncbi.nlm.nih.gov/17571925/)

Martínez-Fernández, R., Gasca-Salas, C., Sánchez-Ferro, Á., & Obeso, J. Á. (2016). Actualización en la enfermedad de Parkinson. *Revista Médica Clínica Las Condes*, 27(3), 363-379.

Marín, D. S., Carmona, H., Ibarra, M., & Gámez, M. (2018). Enfermedad de Parkinson: fisiopatología, diagnóstico y tratamiento. *Revista de La Universidad Industrial de Santander. Salud*, 50(1), 79-92.

Konno, T., Ross, O. A., Puschmann, A., Dickson, D. W. y Wszolek, Z. K. (2016). Enfermedad de Parkinson autosómica dominante causada por duplicaciones de SNCA. *Parkinsonismo y trastornos relacionados*, 22, S1-S6.

Wikipedia contributors. (2025, marzo 28). *Alfa-sinucleína*. Wikipedia.

Trinh, J., Zeldenrust, F.M.J., Huang, J., Kasten, M., Schaaake, S., Petkovic, S., Madoev, H., Grünewald, A., Almuammar, S., König, I.R., Lill, C.M., Lohmann, K., Klein, C. and Marras, C. (2018), Genotype-phenotype relations for the Parkinson's disease genes SNCA, LRRK2, VPS35: MDSGene systematic review. *Mov Disord.*, 33: 1857-1870. <https://doi.org/10.1002/mds.27527>

Rodríguez-Martínez, C. E., Sossa-Briceño, M. P., Payan-Gomez, C., Riaño-Moreno, J., Amador-Munoz, D. P., & Ramírez-Clavijo, S. (2019). Análisis de redes de coexpresión identifica posibles genes clave en el envejecimiento de la corteza prefrontal humana.

Jahabardeen A, S N, J N, V C. A Review on the Role of SNCA Gene in Neurodegenerative Diseases. *Cureus*. 2024 Sep 15;16(9):e69450. doi: 10.7759/cureus.69450. PMID: 39411638; PMCID: PMC11476641

Watanabe H, Shima S, Kawabata K, Mizutani Y, Ueda A, Ito M. Brain network and energy imbalance in Parkinson's disease: linking ATP reduction and α -synuclein pathology. *Front Mol Neurosci*. 2025 Jan 22;17:1507033. doi: 10.3389/fnmol.2024.1507033. PMID: 39911281; PMCID: PMC11794324.

Martijn P. van den Heuvel, Olaf Sporns, Network hubs in the human brain, *Trends in Cognitive Sciences*, Volume 17, Issue 12, 2013, Pages 683-696, ISSN 1364-6613. <https://doi.org/10.1016/j.tics.2013.09.012>.

Akgüller, Ö., Balcı, M. A., & Cioca, G. (2024). Functional Brain Network Disruptions in Parkinson's Disease: Insights from Information Theory and Machine Learning. *Diagnostics*, 14(23), 2728. <https://doi.org/10.3390/diagnostics14232728>

Sang, L., Zhang, J., Wang, L., Zhang, J., Zhang, Y., Li, P., ... & Qiu, M. (2015). Alteration of brain functional networks in early-stage Parkinson's disease: a resting-state fMRI study. *PLoS one*, 10(10), e0141815.

Maraganore, D. M. (2011). Rationale for Therapeutic Silencing of Alpha-Synuclein in Parkinson's Disease. *Journal of Movement Disorders*, 4(1), 1–7.

Henrich, M.T., Oertel, W.H., Surmeier, D.J. *et al.* Mitochondrial dysfunction in Parkinson's disease – a key disease hallmark with therapeutic potential. *Mol Neurodegeneration* 18, 83 (2023). <https://doi.org/10.1186/s13024-023-00676-7>

Zou, L., Tian, Y., & Zhang, Z. (2021). Dysfunction of synaptic vesicle endocytosis in Parkinson's disease. *Frontiers in Integrative Neuroscience*, 15, Article 619160. <https://doi.org/10.3389/fnint.2021.619160>

H. Murali, Mahadevan, A. Hashemiaghdam, G. Ashrafi, A. B. Harbauer, Mitochondria in Neuronal Health: From Energy Metabolism to Parkinson's Disease. *Adv. Biology* 2021, 5, 2100663. <https://doi.org/10.1002/adbi.202100663>

Soukup, S.-F., Vanhauwaert, R., & Verstreken, P. (2018). Parkinson's disease: Convergence on synaptic homeostasis. *The EMBO Journal*, 37(15), e98960. <https://doi.org/10.15252/embj.201898960>

Calabresi, P., Mechelli, A., Natale, G. *et al.* Alpha-synuclein in Parkinson's disease and other synucleinopathies: from overt neurodegeneration back to early synaptic dysfunction. *Cell Death Dis* 14, 176 (2023). <https://doi.org/10.1038/s41419-023-05672-9>

Zhang, B. & Horvath, S. (2005). “A General Framework for Weighted Gene Co-Expression Network Analysis (WGCNA).” *Statistical Applications in Genetics and Molecular Biology*, 4(1).

Love, M.I., Huber, W., Anders, S. (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” *Genome Biology*, 15(12):550.

Shannon, P. *et al.* (2003). “Cytoscape: a software environment for integrated models of biomolecular interaction networks.” *Genome Research*, 13(11): 2498–2504.

Szklarczyk, D. *et al.* (2021). “STRING v11: Protein–protein association networks with increased coverage.” *Nucleic Acids Research*, 49(D1): D605–D612.